

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:35:27 ; Search time 60 seconds
(without alignments)
3475.334 Million cell updates/sec

Title: US-10-049-957-4

Perfect score: 3936

Sequence: 1 MRGPGCALWLLALRTVLGG.....APLPLLPALAAARLLPPAL 738

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3936	100.0	738	2 AAR47899	Human mel
2	3936	100.0	738	4 AAB62881	Membrane
3	3936	100.0	738	5 AAU78363	Cell diff
4	3918	99.5	738	1 AAP70382	Sequence
5	3834	97.4	718	4 AAB73169	Human MTF
6	3441	87.4	736	4 AAB62880	Membrane
7	3441	87.4	736	5 AAU78362	Cell diff
8	3300	83.8	738	4 AAB62882	Membrane
9	3300	83.8	738	4 AAE06668	Mouse p97
10	1850	47.0	502	2 AAU73605	Amino aci
11	1850	47.0	502	2 AAU73616	Amino aci
12	1850	47.0	502	2 AAU73611	Amino ter
13	1850	47.0	502	2 AAU73611	CTLA4/p97
14	1850	47.0	502	2 AAU73611	CTLA4/p97
15	1850	47.0	502	2 AAU73611	CTLA4/p97
16	1295	32.9	708	2 AAR71037	Draculin.
17	1271.5	32.3	698	2 AAR12499	Human tra
18	1269.5	32.3	698	2 AAR66492	Human tra
19	1269.5	32.3	698	2 AAU74354	Serotrans
20	1269.5	32.3	698	3 AAU50717	Human ser
21	1269.5	32.3	698	6 ABP72819	Human tra
22	1269.5	32.3	698	7 ADD45282	Human Pro
23	1264.5	32.1	1410	2 AAU07622	LDLR/TF c
24	1264.5	32.1	1410	6 ABU04139	Human exp
25	1263.5	32.1	679	6 ABP72820	Human tra

26	1262.5	32.1	1074	2 AAW07621	Aaw07621 LDLR/TF c
27	1261	32.0	705	2 AAR22423	Aar22423 Human lac
28	1261	32.0	705	2 AAY31152	Aay31152 Human lac
29	1260.5	32.0	708	2 AAR22424	Aar22424 Human lac
30	1260.5	32.0	708	2 AAY31153	Aay31153 Human lac
31	1257.5	31.9	711	2 AAR08033	Aar08033 Human lac
32	1255.5	31.9	711	2 AAR43653	Lactoferr
33	1255.5	31.9	711	2 AAW09342	Aaw09342 Human lac
34	1255.5	31.9	711	2 AAW57317	Aaw57317 Human lac
35	1255.5	31.9	711	2 AAW86021	Aaw86021 Human lac
36	1255.5	31.9	711	3 AAB08182	Aab08182 Amino aci
37	1255.5	31.9	711	4 AAE02341	Aae02341 Human lac
38	1254	31.9	690	5 AAB080724	Abg080724 Codon opt
39	1254	31.9	690	5 AAE27884	Aae27884 Human cod
40	1252	31.8	687	4 AAG77911	Aag77911 Human lac
41	1252	31.8	688	4 AAG77910	Aag77910 Human lac
42	1252	31.8	689	4 AAG77909	Aag77909 Human lac
43	1252	31.8	690	4 AAG77908	Aag77908 Human lac
44	1252	31.8	692	3 AAY58733	Aay58733 Human lac
45	1252	31.8	692	4 AAB97382	Aab97382 Human lac

ALIGNMENTS

RESULT 1
AAR47899
ID AAR47899 standard; protein; 738 AA.

AC AAR47899;

DT 25-MAR-2003 (revised)

DT 18-AUG-1994 (revised)

DT 20-JUL-1994 (first entry)

XX Human melanoma-associated antigen p97.

XX Human melanoma-associated antigen p97; Melanotransferrin; GPI;

KW glycosyl-phosphatidylinositol anchor; truncated; soluble;

KW blood-brain barrier; iron binding; Alzheimer's disease; iron metabolism;

KW disorder; haemochromatosis; ischaemic tissue damage; heart disease;

XX skin cancer; brain tumour.

OS Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..19

FT Protein 20..738

FT /label= p97

XX WO9401463-A1.

PD 20-JAN-1994.

PF 09-JUL-1993; 93WO-CA0000272.

PR 10-JUL-1992; 92US-00912291.

XX (UYBR-) UNIV BRITISH COLUMBIA.

XX Jefferies WA, McGeer PL, Rothenberger S, Food MR, Yamada T;

XX WPI; 1994-034993/04.

XX N-PSDB; AAQ55768.

XX New GPI-anchored and soluble forms of p97 - for treating disorders of iron metabolism, delivering drugs to the brain and for diagnosis and treatment of Alzheimer's disease.

XX Disclosure; Page 102-106; 166pp; English.

XX The melanoma-associated antigen p97 has been found to be a GPI-anchored protein expressed on the cell surface which is able to bind iron. A

CC soluble form of p97 can be isolated from the aqueous phase after Triton-X
 CC -114 phase separation. The soluble (hydrophilic) form of p97 does not
 CC contain ethanolamine and it has a slower rate of transport than GPI-
 CC anchored p97. The role of p97 in iron-transport suggests a use in
 CC modulating iron uptake by cells; p97, its agonists, antagonists and
 CC stimulants may be useful in treatments of conditions where iron-metabolism
 CC is disturbed, e.g. haemochromatosis and skin cancer. Expression of p97
 CC (and transferrin receptors) by reactive microglial cells associated with
 CC senile plaques in Alzheimer's disease (AD) provides a means of diagnosing
 CC AD. (Updated on 25-MAR-2003 to correct PN field.)
 XX
 XX
 SQ Sequence 738 AA;

Query Match 100.0%; Score 3936; DB 2; Length 738;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 738; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRGPGALWLLALRTVLGGMVVRWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTS 60
 DB 1 MRGPGALWLLALRTVLGGMVVRWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTS 60

QY 61 DHCQVLIQAQADAITLDGGAIYEAAGKEHGLKPVVGVYDQEVGTSYYAVAVRRSSHT 120
 DB 61 DHCQVLIQAQADAITLDGGAIYEAAGKEHGLKPVVGVYDQEVGTSYYAVAVRRSSHT 120

QY 121 IDTLKGVSCHTGINRTVGMNVPVGYLVSGBLSVMGCDVLKAVSDYFGGSCVPAGETS 180
 DB 121 IDTLKGVSCHTGINRTVGMNVPVGYLVSGBLSVMGCDVLKAVSDYFGGSCVPAGETS 180

QY 181 YSESLCRLCRGSSGEGVCDKSPLEYYDYSGAFCLAEAGADVAFVXHSVTLENTDGT 240
 DB 181 YSESLCRLCRGSSGEGVCDKSPLEYYDYSGAFCLAEAGADVAFVXHSVTLENTDGT 240

QY 241 LPSWGQALLSQDFELLCCRDGSRADVTWQRCHLARVPAHAVVVRADTDGGLIFRLNEGQ 300
 DB 241 LPSWGQALLSQDFELLCCRDGSRADVTWQRCHLARVPAHAVVVRADTDGGLIFRLNEGQ 300

QY 301 RLFSHEGSSQFMSSEAYGQKDLLPKDSTSELVPIATQTYEAWLGHEYLHANKGLLCDPN 360
 DB 301 RLFSHEGSSQFMSSEAYGQKDLLPKDSTSELVPIATQTYEAWLGHEYLHANKGLLCDPN 360

QY 361 RLPPYLRCVLTSTPEIQKGDMAVAFRRQLKPEIQCVSAKSPQHMERIOAEQVDVTL 420
 DB 361 RLPPYLRCVLTSTPEIQKGDMAVAFRRQLKPEIQCVSAKSPQHMERIOAEQVDVTL 420

QY 421 SGEDIYTAGKYGKGLVPAAGEHAPEDSSNSYYVAVVRRDSSHAFTLDLGRKRSCHAGF 480
 DB 421 SGEDIYTAGKYGKGLVPAAGEHAPEDSSNSYYVAVVRRDSSHAFTLDLGRKRSCHAGF 480

QY 481 GSPAGWDVPVGLIQRGFIRPKDCDVLTAVERSEFFNASCVPVNNPKYPSLLCALCVGDEQ 540
 DB 481 GSPAGWDVPVGLIQRGFIRPKDCDVLTAVERSEFFNASCVPVNNPKYPSLLCALCVGDEQ 540

QY 541 GRNKCUGNSQERYGVRGAFCLVENAGDVAFRHTTVPDNTNGHNSPAAELRSEDYE 600
 DB 541 GRNKCUGNSQERYGVRGAFCLVENAGDVAFRHTTVPDNTNGHNSPAAELRSEDYE 600

QY 601 LLCPNGARAEVSQFAACNLQAIPPHAVMVRPDTNIFTYVGLLDKQADLFGDDHNNKGFNM 660
 DB 601 LLCPNGARAEVSQFAACNLQAIPPHAVMVRPDTNIFTYVGLLDKQADLFGDDHNNKGFNM 660

QY 661 FDSNNYHGQDLLFKDQATVRAVPVGEKTYRGWLGLDYVAALGMSQOCSGAAAPAGAP 720
 DB 661 FDSNNYHGQDLLFKDQATVRAVPVGEKTYRGWLGLDYVAALGMSQOCSGAAAPAGAP 720

QY 721 LPLPLLALAAALLPPAL 738
 DB 721 LPLPLLALAAALLPPAL 738

RESULT 2
 AAB62881
 ID AAB62881 standard; protein; 738 AA.

XX AAB62881;
 XX 21-MAY-2001 (first entry)
 DT
 XX Membrane bound transferrin like protein amino acid sequence.
 DE
 XX Chondrogenesis promoter; membrane-bound transferrin-like protein; MTF;
 KW Chondrogenesis regulator; Mtf activator; bone metabolism; human;
 KW chondral differentiation inhibitor; bone disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200113951-A1.
 XX
 PD 01-MAR-2001.
 XX
 PF 21-AUG-2000; 2000WO-JP005590.
 XX
 PR 19-AUG-1999; 99JP-00232966.
 XX
 PA (CHUS) CHUGAI SEIYAKU KK.
 XX
 PI Kato Y, Fujimoto K;
 XX
 XX WPI; 2001-218409/22.
 DR N-PSDB; AAP62197.
 XX
 PT Chondrogenesis promoters containing membrane-bound transferrin-like
 PT protein, useful in diagnosis, prevention and treatment of diseases due to
 PT abnormal chondral metabolism and bone metabolism.
 XX
 PS Claim 3; Page 37-41; 57pp; Japanese.
 XX
 CC This invention relates to chondrogenesis promoters containing a membrane-
 CC bound transferrin-like protein (MTf). Chondrogenesis promoters,
 CC chondrogenesis regulators, Mtf activators, Mtf antagonist-containing
 CC chondral differentiation inhibitors are useful in diagnosis, prevention
 CC and treatment of diseases due to abnormal chondral metabolism and bone
 CC metabolism e.g. bone diseases. The present sequence represents the amino
 CC acid sequence of human MTF
 XX
 SQ Sequence 738 AA;

Query Match 100.0%; Score 3936; DB 4; Length 738;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 738; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRGPGALWLLALRTVLGGMVVRWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTS 60
 DB 1 MRGPGALWLLALRTVLGGMVVRWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTS 60

QY 61 DHCQVLIQAQADAITLDGGAIYEAAGKEHGLKPVVGVYDQEVGTSYYAVAVRRSSHT 120
 DB 61 DHCQVLIQAQADAITLDGGAIYEAAGKEHGLKPVVGVYDQEVGTSYYAVAVRRSSHT 120

QY 121 IDTLKGVSCHTGINRTVGMNVPVGYLVSGBLSVMGCDVLKAVSDYFGGSCVPAGETS 180
 DB 121 IDTLKGVSCHTGINRTVGMNVPVGYLVSGBLSVMGCDVLKAVSDYFGGSCVPAGETS 180

QY 181 YSESLCRLCRGSSGEGVCDKSPLEYYDYSGAFCLAEAGADVAFVXHSVTLENTDGT 240
 DB 181 YSESLCRLCRGSSGEGVCDKSPLEYYDYSGAFCLAEAGADVAFVXHSVTLENTDGT 240

QY 241 LPSWGQALLSQDFELLCCRDGSRADVTWQRCHLARVPAHAVVVRADTDGGLIFRLNEGQ 300
 DB 241 LPSWGQALLSQDFELLCCRDGSRADVTWQRCHLARVPAHAVVVRADTDGGLIFRLNEGQ 300

QY 301 RLFSHEGSSQFMSSEAYGQKDLLPKDSTSELVPIATQTYEAWLGHEYLHANKGLLCDPN 360
 DB 301 RLFSHEGSSQFMSSEAYGQKDLLPKDSTSELVPIATQTYEAWLGHEYLHANKGLLCDPN 360

QY 361 RLPPYLRCVLTSTPEIQKGDMAVAFRRQLKPEIQCVSAKSPQHMERIOAEQVDVTL 420

Db 361 RLPPYLRWCVLSTPEIQKCGDMAVAFRRQRLKPEIQCSAKSPQHCHMERIQAEQVDAVTL 420
 QY 421 SGEDIYTAGKKYGLVPAAGHYAPEDSSNSYYVAVVRRDSSHAFTLDELGRKRSCHAGF 480
 Db 421 SGEDIYTAGKKYGLVPAAGHYAPEDSSNSYYVAVVRRDSSHAFTLDELGRKRSCHAGF 480
 QY 481 GSPAGWDVPVGGALIQGFIRPKDCDVLTAVERSEFNASCVPVNNPKNYPSSLCALCVGDEQ 540
 Db 481 GSPAGWDVPVGGALIQGFIRPKDCDVLTAVERSEFNASCVPVNNPKNYPSSLCALCVGDEQ 540
 QY 541 GRNKCVCNSQERYGYRGAFRCCLVENAGDVAFVRHTTVPDNTNGHNSPWAELRSEDYE 600
 Db 541 GRNKCVCNSQERYGYRGAFRCCLVENAGDVAFVRHTTVPDNTNGHNSPWAELRSEDYE 600
 QY 601 LLCPCNGARAEVSQFAACNLAIQIPPHAVMVRPDTNITFTVYGLDLDKAODLFGDDHKNKGF 660
 Db 601 LLCPCNGARAEVSQFAACNLAIQIPPHAVMVRPDTNITFTVYGLDLDKAODLFGDDHKNKGF 660
 QY 661 FDSNNYHGQDLLFKDQATVRAVPVGEKTYRGWLGLDYVAALGCMSSQCSGAAAPAPGAP 720
 Db 661 FDSNNYHGQDLLFKDQATVRAVPVGEKTYRGWLGLDYVAALGCMSSQCSGAAAPAPGAP 720
 QY 721 LLPLLLPALAARLLPPAL 738
 Db 721 LLPLLLPALAARLLPPAL 738

RESULT 3

AAU78363
 ID AAU78363 standard; protein; 738 AA..

XX AAU78363;

XX 18-JUN-2002 (first entry)

DT Cell differentiation stimulator associated protein #4.

XX Cartilage cell differentiation stimulator; osteopathic;
 XX Membrane-bound transferrin-like protein; Mtf-BP; concanavalin A; ConA;
 KW membrane bound type transferrin-like protein; Mtf; cartilage disorder;
 KW bone metabolism disease; cell differentiation; cell growth;
 KW extracellular matrix related disease; human.

XX Homo sapiens.

OS JP2002020311-A.

XX 23-JAN-2002.

XX 07-JUL-2000; 2000JP-00206566.

XX 07-JUL-2000; 2000JP-00206566.

XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

XX WPI; 2002-287405/33.

XX N-PSDB; ABK12569.

XX A cartilage cell differentiation stimulator useful in the diagnosis of
 PT biophylaxis, cell differentiation, cell growth and construction of
 PT extracellular matrix related diseases.
 XX Disclosure; Page 14-16; 17pp; Japanese.

XX The invention describes a cartilage cell differentiation stimulator
 CC (containing a membrane-bound transferrin-like protein (Mtf-BP) and a
 CC membrane bound type transferrin-like protein (Mtf)) and an animal-derived
 CC concanavalin-like drug. The cartilage differentiation stimulator can be
 CC used in diagnosis, prevention and treatment of cartilage and bone
 CC metabolism diseases. They can also be used for diagnosing biophylaxis,
 CC cell differentiation, cell growth and construction of extracellular
 CC matrix related diseases. Mtf-BP strongly stimulates differentiation of

CC cartilage cells and exhibits similar action mechanism with that of plant
 CC derived ConA. This is the amino acid sequence of a cartilage cell
 CC differentiation stimulator associated polypeptide described in the
 CC invention
 XX Sequence 738 AA;

Query Match 100.0%; Score 3936; DB 5; Length 738;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 738; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRGPSALWLLALRTVLGCMVEVRWCATSDPQHKCGNMSEAFREAGIOPSLLCVRGTS 60
 Db 1 MRGPSALWLLALRTVLGCMVEVRWCATSDPQHKCGNMSEAFREAGIOPSLLCVRGTS 60
 QY 61 DHCVCQIIAAQAEADAITLDGGAIIYAGKEHGLKPVVGEVYDQEVGTGYVAVVRRSSHVT 120
 Db 61 DHCVCQIIAAQAEADAITLDGGAIIYAGKEHGLKPVVGEVYDQEVGTGYVAVVRRSSHVT 120
 QY 121 IDTLKGKVSCHTGINRTVGNNVPVGYLVBESGRLSVMGCDVLKAVSDYFGSCVPGAGETS 180
 Db 121 IDTLKGKVSCHTGINRTVGNNVPVGYLVBESGRLSVMGCDVLKAVSDYFGSCVPGAGETS 180
 QY 181 YSESLCRLCRGDSGEGVCDKSPLEYYDYSAGPRCLAEGAGDVAFVKISTVLENTDGT 240
 Db 181 YSESLCRLCRGDSGEGVCDKSPLEYYDYSAGPRCLAEGAGDVAFVKISTVLENTDGT 240
 QY 241 LPSWGOALLSQDFELLCRDGRADVTWEQCHLARPAPAHAVVVRADTDGGLIFRLNNEG 300
 Db 241 LPSWGOALLSQDFELLCRDGRADVTWEQCHLARPAPAHAVVVRADTDGGLIFRLNNEG 300
 QY 301 RLFSHEGSSQFMSSEAYGQKDLLFKDSTSELVPIATQTYEAWLHGYLHANKGLLCDPN 360
 Db 301 RLFSHEGSSQFMSSEAYGQKDLLFKDSTSELVPIATQTYEAWLHGYLHANKGLLCDPN 360
 QY 361 RLPPYLRWCVLSTPEIQKCGDMAVAFRRQRLKPEIQCSAKSPQHCHMERIQAEQVDAVTL 420
 Db 361 RLPPYLRWCVLSTPEIQKCGDMAVAFRRQRLKPEIQCSAKSPQHCHMERIQAEQVDAVTL 420
 QY 421 SGEDIYTAGKKYGLVPAAGHYAPEDSSNSYYVAVVRRDSSHAFTLDELGRKRSCHAGF 480
 Db 421 SGEDIYTAGKKYGLVPAAGHYAPEDSSNSYYVAVVRRDSSHAFTLDELGRKRSCHAGF 480
 QY 481 GSPAGWDVPVGGALIQGFIRPKDCDVLTAVERSEFNASCVPVNNPKNYPSSLCALCVGDEQ 540
 Db 481 GSPAGWDVPVGGALIQGFIRPKDCDVLTAVERSEFNASCVPVNNPKNYPSSLCALCVGDEQ 540
 QY 541 GRNKCVCNSQERYGYRGAFRCCLVENAGDVAFVRHTTVPDNTNGHNSPWAELRSEDYE 600
 Db 541 GRNKCVCNSQERYGYRGAFRCCLVENAGDVAFVRHTTVPDNTNGHNSPWAELRSEDYE 600
 QY 601 LLCPCNGARAEVSQFAACNLAIQIPPHAVMVRPDTNITFTVYGLDLDKAODLFGDDHKNKGF 660
 Db 601 LLCPCNGARAEVSQFAACNLAIQIPPHAVMVRPDTNITFTVYGLDLDKAODLFGDDHKNKGF 660
 QY 661 FDSNNYHGQDLLFKDQATVRAVPVGEKTYRGWLGLDYVAALGCMSSQCSGAAAPAPGAP 720
 Db 661 FDSNNYHGQDLLFKDQATVRAVPVGEKTYRGWLGLDYVAALGCMSSQCSGAAAPAPGAP 720
 QY 721 LLPLLLPALAARLLPPAL 738
 Db 721 LLPLLLPALAARLLPPAL 738

RESULT 4

AAU70382
 ID AAP70382 standard; protein; 738 AA.

XX AAP70382;

XX 25-MAR-2003 (revised)

DT 01-JUL-2002 (revised)

DT 19-FEB-1991 (first entry)

QV	21	MEVRWCATSDPEQHKCGNMSAFREAGIQPSLLVCRGTSADHCVQLIAQAEADAILDGG	80
		Best local similarity 100.0%, E-Val: 0.00e+00	
		Mismatches 0; Indels 0; Gaps 0;	
		Matches 718; Conservative 0;	


```

Db      1 MEVRWCATSPQEHKCNMSEAFREAGIQPSLLCVRGTSADHCVQIAQEAADATLDGG 60
Qy      81 AIYEAGKEHGLKPVGEVYQDEVTGYTAAVAVVRRSSHVTIDTLKGKSCHTGINTVGV 140
Db      61 AIYEAGKEHGLKPVGEVYQDEVTGYTAAVAVVRRSSHVTIDTLKGKSCHTGINTVGV 120
Qy     141 NVPVGLVSGRLSVMGCDLVKAVSDYFGGSCVPGAGETSYSSSLCRLCRGDSGEGVCD 200
Db     121 NVPVGLVSGRLSVMGCDLVKAVSDYFGGSCVPGAGETSYSSSLCRLCRGDSGEGVCD 180
Qy     201 KSPLERYDYGAFRCLEAGAGDVAFKHSTVLENTDGTLPWGQALLSODPELLCRDG 260
Db     181 KSPLERYDYGAFRCLEAGAGDVAFKHSTVLENTDGTLPWGQALLSODPELLCRDG 240
Qy     261 SRADVTEWROCHLARVPAHVVVRAADTDGGLIFRLNEGRLSHGSSQFMSSEAYGQ 320
Db     241 SRADVTEWROCHLARVPAHVVVRAADTDGGLIFRLNEGRLSHGSSQFMSSEAYGQ 300
Qy     321 KOLLFKDSTSELVPIATQTYEAWLGHYELHAMKGLLCDPNRLPPYLRCWCVLSTPEIKCG 380
Db     301 KOLLFKDSTSELVPIATQTYEAWLGHYELHAMKGLLCDPNRLPPYLRCWCVLSTPEIKCG 360
Qy     381 DMVAFRRORLKPEIOCVSAKSPHOCMERIOAEQVDAVTLTSGEDIYTAGKKYGLVPAAGE 440
Db     361 DMVAFRRORLKPEIOCVSAKSPHOCMERIOAEQVDAVTLTSGEDIYTAGKKYGLVPAAGE 420
Qy     441 HYAPESSSYVAVVVRDSSHAFTLDELGRKRSCHAGFGSPAGWDVPVGLIQRGFR 500
Db     421 HYAPESSSYVAVVVRDSSHAFTLDELGRKRSCHAGFGSPAGWDVPVGLIQRGFR 480
Qy     501 PKDCDLVLTAVSEFFNASCVPNNKPNPSSLCALCVGDEGRNKCNGNSQERYYGRGAF 560
Db     481 PKDCDLVLTAVSEFFNASCVPNNKPNPSSLCALCVGDEGRNKCNGNSQERYYGRGAF 540
Qy     561 RCLVENAGDVAFVRHTTVFDNTNGHNSPEWAAELRSEDYELLCPNGARAEVSQFAACNLA 620
Db     541 RCLVENAGDVAFVRHTTVFDNTNGHNSPEWAAELRSEDYELLCPNGARAEVSQFAACNLA 600
Qy     621 QIPPHAVMVRPDNIITVYGLLDKAQDLFGDDHKNKGFKNFDSNNYHGDQLLFKDATVRA 680
Db     601 QIPPHAVMVRPDNIITVYGLLDKAQDLFGDDHKNKGFKNFDSNNYHGDQLLFKDATVRA 660
Qy     681 VPVGKTYRGWGLDYVAALGMSQSCGAAAPAPGAPLLPLLPALAAARLLPPAL 738
Db     661 VPVGKTYRGWGLDYVAALGMSQSCGAAAPAPGAPLLPLLPALAAARLLPPAL 718

```

RESULT 6

```

AAB62880
ID      AAB62880 standard; protein; 736 AA.
AC      AAB62880;
XX
XX
XX      21-MAY-2001 (first entry)
DE      Membrane bound transferrin like protein amino acid sequence.
KW      Chondrogenesis promoter; membrane-bound transferrin-like protein; Mtf;
KW      Chondrogenesis regulator; Mtf activator; bone metabolism; rabbit;
KW      Chondral differentiation inhibitor; bone disease.
XX
XX      Oryctolagus cuniculus.
XX
XX      WO200113951-A1.
XX
XX      01-MAR-2001.
XX
XX      21-AUG-2000; 2000WO-JP005590.
XX
XX      19-AUG-1999; 99JP-00232966.
XX
XX      (CHUS ) CHUGAI SEIYAKU KK.

```

```

XX      Kato Y, Fujimoto K;
PI      WPI; 2001-218409/22.
XX      N-PSDB; AA62196.
DR
XX      Chondrogenesis promoters containing membrane-bound transferrin-like
PT      protein, useful in diagnosis, prevention and treatment of diseases due to
PT      abnormal chondral metabolism and bone metabolism.
XX
XX      Disclosure; Page 32-36; 57pp; Japanese.
XX
XX      This invention relates to chondrogenesis promoters containing a membrane-
CC      bound transferrin-like protein (Mtf). Chondrogenesis promoters,
CC      chondrogenesis regulators, Mtf activators, Mtf antagonist-containing
CC      chondral differentiation inhibitors are useful in diagnosis, prevention
CC      and treatment of diseases due to abnormal chondral metabolism and bone
CC      metabolism e.g. bone diseases. The present sequence represents the amino
CC      acid sequence of rabbit Mtf
XX
XX      Sequence 736 AA;
SQ

```

Query Match 87.4%; Score 3441; DB 4; Length 736;

Best Local Similarity 86.7%; Pred. No. 1.8e-302; Mismatches 50; Indels 2; Gaps 1;

Matches 640; Conservative 46; Mismatches 50; Indels 2; Gaps 1;

```

Qy      1 MRGPSALMLLALRTVLGMEVRWCATSDPEQHKCNMSEAFREAGIQPSLLCVRGTS 60
Db      1 MCRSAAWIFLALRTALGSEVRWCATSEPEQKCEDMSQAFREAGLQPALLCVOGTS 60
Qy     61 DHCQVLIAAQDAITLDGGAIEYAGKEHGLKPVGEVYQDEVTGYTAAVAVVRRSSHVT 120
Db     61 DHCQVLIAAQDAITLDGGAIEYAGKEHGLKPVGEVYQDEVTGYTAAVAVVRRSSHVT 120
Qy     121 IDTLKGVKSCHTGINRTVGMNVPVGYLVESGRLSVMGCDVLKAVSDYFGGSCVPGAGET 180
Db     121 IDTLKGVKSCHTGINRTVGMNVPVGYLVESGRLSVMGCDVLKAVSDYFGGSCVPGAGET 180
Qy     181 YSLSLRLCRGSSGEGVCDKSPLERYDYSGAFRCLEAGAGDVAFVGSTVLENTDGT 240
Db     181 YSLSLRLCRGSSGEGVCDKSPLERYDYSGAFRCLEAGAGDVAFVGSTVLENTDGT 240
Qy     241 LPSWGQALLSQDFELLCLRDGSRADVTWROCHLARVPAHVVVRAADTDGGLIFRLNEGQ 300
Db     241 LPSWGQALLSQDFELLCLRDGSRADVTWROCHLARVPAHVVVRAADTDGGLIFRLNEGQ 300
Qy     301 RLFSHSGSSQFMSSEAYGQKOLLFKDSTSELVPIATQTYEAWLGHYELHAMKGLLCDPN 360
Db     301 RLFSHSGSSQFMSSEAYGQKOLLFKDSTSELVPIATQTYEAWLGHYELHAMKGLLCDPN 360
Qy     361 RLPPYLRCWCVLSTPEIKQKGDMAVAFRRORLKPFIQCVSAKSPHOCMERIOAEQVDAVTL 420
Db     361 RLPPYLRCWCVLSTPEIKQKGDMAVAFRRORLKPFIQCVSAKSPHOCMERIOAEQVDAVTL 420
Qy     421 SGEDIYTAGKKYGLVPAAGEHYAPEDSSNSYVAVVRRDSSHAFTLDELGRKRSCHAG 480
Db     421 SGEDIYTAGKKYGLVPAAGEHYAPEDSSNSYVAVVRRDSSHAFTLDELGRKRSCHAG 480
Qy     481 GSPAGWDVPVGLIQRGFTIRPKDCDLVLTAVSEFFNASCVPNNKPNPSSLCALCVGDEQ 540
Db     481 GSPAGWDVPVGLIQRGFTIRPKDCDLVLTAVSEFFNASCVPNNKPNPSSLCALCVGDEQ 540
Qy     541 GRNKCNGNSQERYYGRGAFRCLEAGAGDVAFVRHTTVFDNTNGHNSPEWAAELRSEDYE 600
Db     541 GRNKCNGNSQERYYGRGAFRCLEAGAGDVAFVRHTTVFDNTNGHNSPEWAAELRSEDYE 600
Qy     601 LLCPNGARAEVSQFAACNLAQIPPHAVMVRPDNIITVYGLLDKAQDLFGDDHKNKGFKN 660
Db     601 LLCPNGARAEVSQFAACNLAQIPPHAVMVRPDNIITVYGLLDKAQDLFGDDHKNKGFKN 660
Qy     661 FDSSSYHGDQLLFKDATVRAVPVGKTYRGWGLDYVAALGMSQSCGAAAPAPGAP 720
Db     661 FDSSSYHGDQLLFKDATVRAVPVGKTYRGWGLDYVAALGMSQSCGAAAPAPGAP 720

```


CC regulating CTLA4 receptor positive T cell interactions with B7 receptor
 CC positive B cells. The method comprises contacting the CTLA4-positive T
 CC cells with monoclonal antibody fragments reactive with CTLA4. This
 CC inhibits (and therefore regulates) interactions between CTLA4-positive T
 CC cells and B7 positive B cells. The method may be used for regulating
 CC CTLA4 receptor positive T cell interactions with B7 receptor positive B
 CC cells. In this way the immune system of an individual can be manipulated
 CC (especially suppressed) for the treatment of autoimmune diseases
 CC (especially lupus erythematosus) and to prevent host-graft and transplant
 CC rejection

XX SQ Sequence 502 AA;

Query Match 47.0%; Score 1850; DB 2; Length 502;
 Best Local Similarity 100.0%; Pred. No. 2.1e-158;
 Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GMEVRWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVOLIAAQEADAITLDG 79
 DB 156 GMEVRWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVOLIAAQEADAITLDG 215
 QY 80 GAIYEAGKEHGLKPVVGEVYDQVGTSYAVAVVRSSSHVTIDTLGVKVSCHTGINRTVG 139
 DB 216 GAIYEAGKEHGLKPVVGEVYDQVGTSYAVAVVRSSSHVTIDTLGVKVSCHTGINRTVG 275
 QY 140 MNVPVGYLVESGRSLVWGCDVLKAVSDYFGGSCVPGAGETSYSESLCRLCRGDSGEGVC 199
 DB 276 MNVPVGYLVESGRSLVWGCDVLKAVSDYFGGSCVPGAGETSYSESLCRLCRGDSGEGVC 335
 QY 200 DKSPLEYYDYSAGFRCLAEAGADVAFVKHSTVLENTDGTLPWSQALLSQDFELLCRD 259
 DB 336 DKSPLEYYDYSAGFRCLAEAGADVAFVKHSTVLENTDGTLPWSQALLSQDFELLCRD 395
 QY 260 GSRADVTWEHQHARVPAHVVVRADTDGGLIFRLNNEGQRLFSHEGSSFQMFSEAYG 319
 DB 396 GSRADVTWEHQHARVPAHVVVRADTDGGLIFRLNNEGQRLFSHEGSSFQMFSEAYG 455
 QY 320 QKDLLEFKDSTSELVPIATQTYEAWLGHEYLHAKMGLLCDPNRLPPYL 366
 DB 456 QKDLLEFKDSTSELVPIATQTYEAWLGHEYLHAKMGLLCDPNRLPPYL 502

RESULT 11

AAW97616
 ID AAW97616 standard; protein; 502 AA.

XX AC AAW97616;

XX DT 14-JUN-1999 (first entry)

XX DE Amino terminal CTLA4-carboxy terminal p97 fusion protein.

XX KW CTLA4 receptor; immunoglobulin; Ig; fusion protein; B7 antigen;
 KW soluble CTLA4; T cell interaction; B7 positive cell; immune disorder;
 KW immune system disease; graft versus host disease; psoriasis;
 KW graft transplant rejection; T cell lymphoma; benign lymphocytic angitis;
 KW autoimmune disease; lupus erythematosus; Grave's disease;
 KW Addison's disease; Crohn's disease; multiple sclerosis;
 KW ulcerative colitis; Sjogren's syndrome; mixed connective tissue disease;
 KW viral proliferation; T cell activation; AIDS; HTLV1.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN US5885579-A.

XX PD 23-MAR-1999.

XX PF 08-JUL-1997; 97US-00889666.

XX PR 27-JUN-1991; 91US-00723617.

XX PR 22-JAN-1993; 93US-00008898.

XX PR 15-APR-1994; 94US-00228208.

PR 18-JAN-1995; 95US-00375390.

XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.

XX PI Damle NK, Kiener PA, Brady W, Ledbetter JA, Linsley PS;

XX DR WPI; 1999-228484/19.

XX PT Human CTLA4 receptor protein - used to regulate T-cell interactions with
 PT B7 positive cells.

XX PS Disclosure; Fig 36; 75pp; English.

XX CC The specification describes a CTLA4 receptor/immunoglobulin (Ig) fusion
 CC protein which is reactive with the B7 antigen. DNA encoding the CTLA4
 CC protein can be used in a host vector system for producing
 CC soluble CTLA4. The CTLA4 fusion protein can be used for regulating T cell
 CC interactions with B7 positive cells. The CTLA4Ig fusion protein can be
 CC used for treating immune system diseases mediated by T cell interactions
 CC with B7 positive cells. The immune system diseases include graft versus
 CC host disease, psoriasis, immune disorders associated with graft
 CC transplant rejection, T cell lymphoma, benign lymphocytic angitis, and
 CC autoimmune diseases such as lupus erythematosus, Grave's disease,
 CC Addison's disease, Crohn's disease, multiple sclerosis, ulcerative
 CC colitis, Sjogren's syndrome, and mixed connective tissue disease. The
 CC fusion protein may also be used to block the proliferation of viruses
 CC dependent on T cell activation, such as the virus that causes AIDS,
 CC HTLV1. The present sequence was created in the course of the invention

XX SQ Sequence 502 AA;

Query Match 47.0%; Score 1850; DB 2; Length 502;
 Best Local Similarity 100.0%; Pred. No. 2.1e-158;
 Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GMEVRWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVOLIAAQEADAITLDG 79
 DB 156 GMEVRWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVOLIAAQEADAITLDG 215
 QY 80 GAIYEAGKEHGLKPVVGEVYDQVGTSYAVAVVRSSSHVTIDTLGVKVSCHTGINRTVG 139
 DB 216 GAIYEAGKEHGLKPVVGEVYDQVGTSYAVAVVRSSSHVTIDTLGVKVSCHTGINRTVG 275

QY 140 MNVPVGYLVESGRSLVWGCDVLKAVSDYFGGSCVPGAGETSYSESLCRLCRGDSGEGVC 199

DB 276 MNVPVGYLVESGRSLVWGCDVLKAVSDYFGGSCVPGAGETSYSESLCRLCRGDSGEGVC 335

QY 200 DKSPLEYYDYSAGFRCLAEAGADVAFVKHSTVLENTDGTLPWSQALLSQDFELLCRD 259

DB 336 DKSPLEYYDYSAGFRCLAEAGADVAFVKHSTVLENTDGTLPWSQALLSQDFELLCRD 395

QY 260 GSRADVTWEHQHARVPAHVVVRADTDGGLIFRLNNEGQRLFSHEGSSFQMFSEAYG 319

DB 396 GSRADVTWEHQHARVPAHVVVRADTDGGLIFRLNNEGQRLFSHEGSSFQMFSEAYG 455

QY 320 QKDLLEFKDSTSELVPIATQTYEAWLGHEYLHAKMGLLCDPNRLPPYL 366

DB 456 QKDLLEFKDSTSELVPIATQTYEAWLGHEYLHAKMGLLCDPNRLPPYL 502

RESULT 12

AAW97611

ID AAW97611 standard; protein; 502 AA.

XX AC AAW97611;

XX DT 26-MAY-1999 (first entry)

XX DE Amino terminal CTLA4-carboxy terminal p97 fusion protein.

XX KW CTLA4 receptor; immunoglobulin; Ig; fusion protein; B7 antigen;

KW soluble CTLA4; T cell interaction; B7 positive cell; immune disorder;

KW immune system disease; graft versus host disease; psoriasis;

KW graft transplant rejection; T cell lymphoma; benign lymphocytic angitis;
 KW autoimmune disease; lupus erythematosus; Grave's disease;
 KW Addison's disease; Crohn's disease; multiple sclerosis;
 KW ulcerative colitis; Sjogren's syndrome; mixed connective tissue disease;
 KW viral proliferation; T cell activation; AIDS; HTLV1.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 XX US5885796-A.
 XX
 XX 23-MAR-1999.
 XX
 XX 05-JUN-1995; 95US-00465078.
 XX
 XX 27-JUN-1991; 91US-00723617.
 XX
 XX 22-JAN-1993; 93US-00008898.
 XX
 XX 15-APR-1994; 94US-00228208.
 XX
 XX 18-JAN-1995; 95US-00375390.
 XX
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 XX Damle NK, Brady W, Ledbetter JA, Linsley PS;
 XX
 XX WPI; 1999-228535/19.
 XX
 XX CTLA4 receptor protein for use in treatment of immune system diseases.
 XX
 XX Disclosure; Fig 36; 75pp; English.
 XX
 XX The specification describes a CTLA4 receptor/immunoglobulin (Ig) fusion
 CC protein which is reactive with the B7 antigen. DNA encoding the CTLA4
 CC protein can be used in a host vector system for producing
 CC soluble CTLA4. The CTLA4 fusion protein can be used for regulating T cell
 CC interactions with B7 positive cells. The CTLA4 fusion protein can be
 CC used for treating immune system diseases mediated by T cell interactions
 CC with B7 positive cells. The immune system diseases include graft versus
 CC host disease, psoriasis, immune disorders associated with graft
 CC transplant rejection, T cell lymphoma, benign lymphocytic angitis, and
 CC autoimmune diseases such as lupus erythematosus, Grave's disease,
 CC Addison's disease, Crohn's disease, multiple sclerosis, ulcerative
 CC colitis, Sjogren's syndrome, and mixed connective tissue disease. The
 CC fusion protein may also be used to block the proliferation of viruses
 CC dependent on T cell activation, such as the virus that causes AIDS.
 CC HTLV1. The present sequence was created in the course of the invention
 XX
 XX Sequence 502 AA;
 XX
 XX Query Match 47.0%; Score 1850; DB 2; Length 502;
 XX Best Local Similarity 100.0%; Pred. No. 2.1e-158;
 XX Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 20 GMEVWCATSDPEQHKCGNMSAFREAGIQPSLLCVRGTSADHCVQLIAAQEADAITLDG 79
 Db 156 GMEVWCATSDPEQHKCGNMSAFREAGIQPSLLCVRGTSADHCVQLIAAQEADAITLDG 215
 QY 80 GAIYAGKEHGLKPVGEYDQVETGTSYAVAVRRSSHTVITDLTKGVKSCHTGINTVVG 139
 Db 216 GAIYAGKEHGLKPVGEYDQVETGTSYAVAVRRSSHTVITDLTKGVKSCHTGINTVVG 275
 QY 140 WNPVGYLVESGRSLVWVGCDVLKAVSDYFGGSCVPGAGTSSYSELCLRCRDSGEGVC 199
 Db 276 WNPVGYLVESGRSLVWVGCDVLKAVSDYFGGSCVPGAGTSSYSELCLRCRDSGEGVC 335
 QY 200 DKSPLEYYDYSGAFCLAEAGDVAFKVHSTVLENTDGTLPNGQALLSODFELLCRD 259
 Db 336 DKSPLEYYDYSGAFCLAEAGDVAFKVHSTVLENTDGTLPNGQALLSODFELLCRD 395
 QY 260 GSRADVTEWROCHLARPVAHVAVRADTDGGLIFRLNBEQRLFSHEGSSFOFMSSEAYG 319
 Db 396 GSRADVTEWROCHLARPVAHVAVRADTDGGLIFRLNBEQRLFSHEGSSFOFMSSEAYG 455
 QY 320 QKDLLFKDSTSELVPIATQTYEAWLGHEYLHANKGLLCDPNRLPPYL 366

Db 456 QKDLLFKDSTSELVPIATQTYEAWLGHEYLHANKGLLCDPNRLPPYL 502
 RESULT 13
 ID AAY41131 standard; protein; 502 AA.
 XX
 XX AAY41131;
 XX
 XX 24-JAN-2000 (first entry)
 XX
 XX CTLA4/p97 fusion protein.
 XX
 XX Monoclonal antibody; MAb; extracellular domain; CTLA4; B7 antigen;
 KW T cell interaction; inflammation; autoimmunity; transplantation; GCHD;
 KW neoplasia; infectious disease; graft versus host disease; psoriasis;
 KW immune disorder; lymphoma; leukemia; autoimmune disease; arthritis;
 KW diabetes mellitus; oncostatin M; fusion protein; p97.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 XX US5977318-A.
 XX
 XX 02-NOV-1999.
 XX
 XX 07-JUN-1995; 95US-00488062.
 XX
 XX 27-JUN-1991; 91US-00723617.
 XX
 XX 22-JAN-1993; 93US-00008898.
 XX
 XX 15-APR-1994; 94US-00228208.
 XX
 XX 18-JAN-1995; 95US-00375390.
 XX
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 XX Kiener PA, Brady W, Damle NK, Linsley PS, Ledbetter JA;
 XX
 XX WPI; 1999-619712/53.
 XX
 XX New anti-CTLA4 monoclonal antibodies, used for treating e.g.
 PT inflammation, autoimmunity, transplant rejection, infectious diseases or
 PT neoplasia.
 XX
 XX Example; Fig 36; 74pp; English.
 XX
 CC The invention provides new monoclonal antibodies (MABs) which bind the
 CC extracellular domain of CTLA4 and prevent the binding of CTLA4 to B7
 CC antigen. The MABs can be used for regulating T cell interactions with B7
 CC positive cells. They can also be used for preventing or reversing
 CC inflammation and for treating autoimmunity, transplantation, infectious
 CC diseases and neoplasia. They can be used for treating diseases e.g. graft
 CC versus host disease (GCHD), psoriasis, immune disorders associated with
 CC graft transplantation rejection, T cell lymphoma, T cell acute
 CC lymphoblastic leukemia, testicular angiocentric R cell lymphoma, benign
 CC lymphocytic angitis, autoimmune diseases such as lupus erythematosus,
 CC Hashimoto's thyroiditis, primary myxedema, Graves disease, pernicious
 CC anemia, autoimmune atrophic gastritis, Addison's disease, insulin
 CC dependent diabetes mellitus, Goodpasture's syndrome, myasthenia gravis,
 CC pemphigus, Crohn's disease, sympathetic ophthalmia, autoimmune uveitis,
 CC multiple sclerosis, autoimmune hemolytic anemia, primary biliary
 CC cirrhosis, idiopathic thrombocytopenia, chronic action hepatitis,
 CC ulcerative colitis, Sjogren's syndrome, rheumatoid arthritis,
 CC polymyositis, scleroderma, and mixed connective tissue disease. They can
 CC also be used for detection, diagnosis, prognosis and monitoring of
 CC diseases. The present sequence represents the CTLA4/p97 fusion protein
 CC containing an amino terminal CTLA4 domain and a p97 carboxy-terminal
 CC domain
 XX
 XX Sequence 502 AA;
 XX
 XX Query Match 47.0%; Score 1850; DB 2; Length 502;
 XX Best Local Similarity 100.0%; Pred. No. 2.1e-158;

Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GMEVRCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVQIIAAQADAITLDG 79
DB 156 GMEVRCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVQIIAAQADAITLDG 215

QY 80 GAIYEAGKEHGLKPVVGEYDQEVGTSYAVAVVRRSSHVTIDTLKGVSCHTGINRTVG 139
DB 216 GAIYEAGKEHGLKPVVGEYDQEVGTSYAVAVVRRSSHVTIDTLKGVSCHTGINRTVG 275

QY 140 WNPVGVYLVESGRLSVMGCDVLKAVSDYFGGSCVPGAGETSYSESLCRICRGDSSGEGVC 199
DB 276 WNPVGVYLVESGRLSVMGCDVLKAVSDYFGGSCVPGAGETSYSESLCRICRGDSSGEGVC 335

QY 200 DKSPLEYYDYSGAFRCLEAGAGDVAFVKHSTVLENTDGTLPWSGQALLSODFELLCRD 259
DB 336 DKSPLEYYDYSGAFRCLEAGAGDVAFVKHSTVLENTDGTLPWSGQALLSODFELLCRD 395

QY 260 GSRADVTWRQCHLARVPAHAVVVRADTDGGLIFRLNNEGQRLFSHEGSSQFMFSSEAYG 319
DB 396 GSRADVTWRQCHLARVPAHAVVVRADTDGGLIFRLNNEGQRLFSHEGSSQFMFSSEAYG 455

QY 320 QKDLLFKDSTSELVPIATQTYEAWLGHEYLHAKMGLLCDPNRLPPYL 366
DB 456 QKDLLFKDSTSELVPIATQTYEAWLGHEYLHAKMGLLCDPNRLPPYL 502

RESULT 14
ID AAW81585
AC AAW81585; standard; protein; 502 AA.
XX
XX
DT 05-FEB-1999 (first entry)
XX
XX
DE CTLA4/p97 fusion protein sequence.
XX
KW CTLA4 receptor; CTLA4-Ig; fusion protein; B7 antigen; hinge; CH2; CH3;
KW extracellular domain; human; immunoglobulin; T cell; immune system;
KW autoimmune disease; cancer; viral infection; p97.
XX
OS Homo sapiens.
OS Unidentified.

XX
XX
FH Key
FT Peptide
FT Protein
FT Protein
FT Domain
FT Domain
FT Protein
FT Protein
FT Domain
FT Domain
XX
XX
PN US5844095-A.
XX
PD 01-DEC-1998.
XX
PF 18-JAN-1995; 95US-00375390.
XX
XX 27-JUN-1991; 91US-00723617.
PR 22-JAN-1993; 93US-00008898.
PR 28-MAY-1993; 93US-00069693.
PR 15-APR-1994; 94US-00228208.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Brady W, Linsley PS, Damle NK, Ledbetter JA;
XX
XX WPI; 1999-044666/04.
XX

PT Fusion protein of CTLA4 and immunoglobulin fragment - for treating immune
XX system disorders.
PS
XX Example 8; Fig 36; 75pp; English.
XX
CC This represents a CTLA4/p97 fusion protein. The invention provides a
CC CTLA4-Ig fusion protein that binds the B7 antigen and has a first amino
CC acid sequence consisting of the extracellular domain of CTLA4 and a
CC second amino acid sequence consisting of the hinge, CH2 and CH3 regions
CC of a human immunoglobulin molecule. The fusion protein inhibits
CC interaction of T cells with B7-positive cells and may be useful for
CC treating immune system diseases, e.g. autoimmune diseases, cancer or
CC viral infections. The present sequence is an example of such a fusion
CC protein and contains an amino-terminal CTLA4 domain and a p97 carboxy-
XX terminal domain
SQ Sequence 502 AA;

Query Match 47.0%; Score 1850; DB 2; Length 502;
Best Local Similarity 100.0%; Pred. No. 2.1e-158;
Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GMEVRCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVQIIAAQADAITLDG 79
DB 156 GMEVRCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVQIIAAQADAITLDG 215

QY 80 GAIYEAGKEHGLKPVVGEYDQEVGTSYAVAVVRRSSHVTIDTLKGVSCHTGINRTVG 139
DB 216 GAIYEAGKEHGLKPVVGEYDQEVGTSYAVAVVRRSSHVTIDTLKGVSCHTGINRTVG 275

QY 140 WNPVGVYLVESGRLSVMGCDVLKAVSDYFGGSCVPGAGETSYSESLCRICRGDSSGEGVC 199
DB 276 WNPVGVYLVESGRLSVMGCDVLKAVSDYFGGSCVPGAGETSYSESLCRICRGDSSGEGVC 335

QY 200 DKSPLEYYDYSGAFRCLEAGAGDVAFVKHSTVLENTDGTLPWSGQALLSODFELLCRD 259
DB 336 DKSPLEYYDYSGAFRCLEAGAGDVAFVKHSTVLENTDGTLPWSGQALLSODFELLCRD 395

QY 260 GSRADVTWRQCHLARVPAHAVVVRADTDGGLIFRLNNEGQRLFSHEGSSQFMFSSEAYG 319
DB 396 GSRADVTWRQCHLARVPAHAVVVRADTDGGLIFRLNNEGQRLFSHEGSSQFMFSSEAYG 455

QY 320 QKDLLFKDSTSELVPIATQTYEAWLGHEYLHAKMGLLCDPNRLPPYL 366
DB 456 QKDLLFKDSTSELVPIATQTYEAWLGHEYLHAKMGLLCDPNRLPPYL 502

RESULT 15
AAW87561
ID AAW87561 standard; protein; 502 AA.
XX
XX AAW87561;
XX
DT 01-MAR-1999 (first entry)
XX
XX CTLA4-p97 fusion protein.
XX
XX CD28; B7; fusion protein; hinge CH2; CH3; human IgC-gamma1; CTLA4;
KW CTLA4 receptor; ligand; regulation; T-cell interaction; B7-positive cell;
KW immune system disease.
XX
OS Synthetic.
OS Homo sapiens.
XX
XX
FH Key
FT Peptide
FT Protein
FT Protein
FT Protein
FT Protein
XX
XX US5851795-A.

Location/Qualifiers
1. .15 /note= "signal peptide"
16. .155 /note= "CTLA4 protein"
155. .502 /note= "p97 protein"

XX 22-DEC-1998.
XX
XX
XX 02-JUN-1995; 95US-00459818.
XX
XX 27-JUN-1991; 91US-00723617.
XX 22-JAN-1993; 93US-00088898.
XX 15-APR-1994; 94US-00228208.
XX 18-JAN-1995; 95US-00375390.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Ledbetter JA, Brady W, Damle NK, Kiener PA, Linsley PS;
XX WPI; 1999-080402/07.
XX
XX Soluble CTLA4 protein that binds B7 antigen of activated B cells - and
XX fusion proteins useful for regulating T-cell interactions with B cells.
XX
XX Claim 16; Fig 36; 75pp; English.
XX
XX The present sequence represents a CTLA4-p97 fusion protein. CTLA4 is a
XX receptor protein having a high degree of homology with CD28. The CTLA4
XX receptor is identified as a ligand for the B7 antigen. The CTLA4 protein
XX can also be used to construct a fusion protein of CTLA4 and human
XX immunoglobulin (Ig)-gamma1. The CTLA4 protein was fused to the hinge CH2
XX and CH3 regions of human Ig-gamma1. Soluble CTLA4-Ig fusion proteins can
XX be used to regulate T-cell interactions with B7-positive cells and to
XX treat immune system diseases mediated by such interactions
XX
XX Sequence 502 AA;
SQ

Query Match 47.0%; Score 1850; DB 2; Length 502;
Best Local Similarity 100.0%; Pred. No. 2.1e-158;
Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 GNEVWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVQLIAAQEADAITLDG 79
Db 156 GNEVWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVQLIAAQEADAITLDG 215

QY 80 GAIYEAGKEHGLKPVVGEVYDQEVGTSYYAVAVVRRSSHVTITDLKGVKSCHTGINRTVG 139
Db 216 GAIYEAGKEHGLKPVVGEVYDQEVGTSYYAVAVVRRSSHVTITDLKGVKSCHTGINRTVG 275

QY 140 WNPVGYLVESGRLSVMGCDVLKAVSDYFGGSCVFCAGETSYSESICRLCRGDSGEGVC 199
Db 276 WNPVGYLVESGRLSVMGCDVLKAVSDYFGGSCVFCAGETSYSESICRLCRGDSGEGVC 335

QY 200 DKSPLERYDYSGAFRCLEAGAGDVAFVKHSTVLENTDKTLPWGOALLSODPELLCRD 259
Db 336 DKSPLERYDYSGAFRCLEAGAGDVAFVKHSTVLENTDKTLPWGOALLSODPELLCRD 395

QY 260 GSRADVTEWROCHLARVPAHAVVVRADTDGGLIFRLNEGQRLFSEHSSFFQFSEAYG 319
Db 396 GSRADVTEWROCHLARVPAHAVVVRADTDGGLIFRLNEGQRLFSEHSSFFQFSEAYG 455

QY 320 QKDLLFKDSTSELVPATQTYEAWLGHYHAMKGLLCDPNRLPPYL 366
Db 456 QKDLLFKDSTSELVPATQTYEAWLGHYHAMKGLLCDPNRLPPYL 502

Search completed: May 14, 2004, 09:41:51
Job time : 64 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:39:47 ; Search time 21 Seconds
(without alignments)
3380.445 Million cell updates/sec

Title: US-10-049-957-4
Perfect score: 3936
Sequence: 1 MRGPSALMILLALRTVLGG.....APLLPLLALAAARLLPPAL 738

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:.*
1: piri:.*
2: piri2:.*
3: piri3:.*
4: piri4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3936	100.0	738	1 TFHUM	melanotransferrin
2	1272.5	32.3	690	2 T11749	transferrin - Atla
3	1269.5	32.3	698	1 TFHUP	transferrin precu
4	1263.5	32.1	706	2 S33761	transferrin precu
5	1260.5	32.0	711	1 TFHUL	lactotransferrin p
6	1258	32.0	708	1 JFC2323	lactoferrin - goat
7	1249	31.7	708	1 TFBOL	lactotransferrin p
8	1228	31.2	694	1 TFRBP	transferrin precu
9	1213.5	30.8	696	1 S01384	transferrin - pig
10	1212	30.8	703	2 A45543	lactoferrin precu
11	1202.5	30.6	704	1 TFC228	carbonic anhydrase
12	1202.5	30.6	705	1 TFCHE	ovotransferrin pre
13	1192.5	30.3	707	1 A28438	lactoferrin precu
14	1160	29.5	695	2 S49163	transferrin precu
15	1128.5	28.7	717	2 S12100	transferrin precu
16	824	20.9	726	2 A47275	transferrin - cock
17	551.5	14.0	681	1 A36500	transferrin precu
18	450	11.4	1274	2 T10729	transferrin-like p
19	408	10.4	215	2 A60166	hemiferrin - bovin
20	408	10.4	216	1 A39684	hemiferrin - rat
21	397.5	10.1	311	2 A28446	transferrin - mous
22	365	9.3	629	2 S68986	transferrin - fles
23	152.5	3.9	87	2 A39725	transferrin - shee
24	145	3.7	119	2 A39426	saxiphilin - bullf
25	121	3.1	1251	2 A57293	latent transferrin
26	121	3.1	1639	1 MMFFB2	laminin gamma-1 ch
27	119	3.0	770	2 T00204	LDL receptor relat
28	114	2.9	1562	2 T17411	polyketide synthas
29	113.5	2.9	770	2 T00203	LDL receptor-relat

30	108	2.7	5369	2	T44807	mycosubtilin synth
31	107	2.7	515	2	T06262	probable protein d
32	106	2.7	1031	2	T43458	hypothetical prote
33	104.5	2.7	513	2	T05974	protein disulfide-
34	102	2.6	660	2	T02755	fanconi anemia com
35	102	2.6	661	2	T42754	hypothetical prote
36	102	2.6	853	2	H70939	probable nirB prot
37	102	2.6	1173	2	B97208	pyruvate ferredoxi
38	101	2.6	1574	2	T13954	MEGF6 protein - ra
39	100.5	2.6	444	2	E82499	hypothetical prote
40	100.5	2.6	1743	2	T26859	hypothetical prote
41	100	2.5	913	2	F82958	DNA polymerase I p
42	100	2.5	1694	2	A83512	hypothetical prote
43	100	2.5	8563	2	T30226	polyketide synthas
44	99	2.5	929	2	A84379	alanyl-tRNA synthe
45	98.5	2.5	655	1	A46688	hepatocyte growth

ALIGNMENTS

RESULT 1
TFHUM

N;Alternate names: melanoma-associated antigen gp95/p97
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 28-Jan-2000
C;Accession: A23814; A60925
R;Rose, T.M.; Plozman, G.D.; Teplow, D.B.; Dreyer, W.J.; Hellstrom, K.E.; Brown, J.P.
Proc. Natl. Acad. Sci. U.S.A. 83, 1261-1265, 1986
A;Title: Primary structure of the human melanoma-associated antigen p97 (melanotransfer
A;Reference number: A23814; MUID:86149285; PMID:2419904
A;Accession: A23814
A;Molecule type: mRNA
A;Residues: 1-738 <ROS>
A;Cross-references: EMBL:M12154; NID:gl89515; PIDN:AA59992.1; PID:gl89518
A;Experimental source: melanoma
R;Furukawa, K.S.; Furukawa, K.; Real, F.X.; Old, L.J.; Lloyd, K.O.
J. Exp. Med. 169, 585-590, 1989
A;Title: A unique antigenic epitope of human melanoma is carried on the common melanoma
A;Reference number: A60925; MUID:89094252; PMID:2463331
A;Accession: A60925
A;Molecule type: protein
A;Residues: 20-25,'X',27-28,'X',30 <FUR>
C;Comment: This protein is found predominantly in human melanomas and in certain fetal
C;Comment: Seven disulfide bonds are predicted in each domain.
C;Genetics:
A;Gene: GDB:MFI2
A;Cross-references: GDB:119387; OMIM:155750
A;Map position: 3q28-3q29
C;Superfamily: transferrin; transferrin repeat homology
C;Keywords: blocked carboxyl end; duplication; glycoprotein; lipoprotein; membrane prot
F;1-19/Domain: signal sequence #status predicted <SIG>
F;18-360/Domain: transferrin repeat homology <TRHi>
F;20-710/Product: melanotransferrin #status predicted <MTF>
F;361-709/Domain: transferrin repeat homology <TRH2>
F;711-738/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;38,135,515/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;710/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature for

Query Match 100.0%; Score 3936; DB 1; Length 738;
Best Local Similarity 100.0%; Pred. No. 7.9e-292;
Matches 738; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MRGPSALMILLALRTVLGGMEVRWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTS	60
Db	1	MRGPSALMILLALRTVLGGMEVRWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTS	60
QY	61	DHCVQLIAAQEADAITLDGGAIEAGKEHGLKPVGVEVDQEVGTGYAVAVVRRSSHVT	120
Db	61	DHCVQLIAAQEADAITLDGGAIEAGKEHGLKPVGVEVDQEVGTGYAVAVVRRSSHVT	120
QY	121	IDTLKGKVSCHTGINRTGNVNPVGYLVESGRLSVMGCDVLKAVSDYFGGSCVPAGETS	180

R;Park, I.; Schaeffer, E.; Sidoli, A.; Baralle, F.E.; Cohen, G.N.; Zakin, M.M.
Proc. Natl. Acad. Sci. U.S.A. 82, 3149-3153, 1985
A;Title: Organization of the human transferrin gene: direct evidence that it originated
A;Reference number: A94044; MUID:85216459; PMID:3858812
A;Accession: A94044
A;Molecule type: DNA
A;Residues: 73-263, 'E', 265-328, 'N', 330-562 <PAR>
A;Cross-references: EMBL:MI1361
R;Adrian, G.S.; Korinek, B.W.; Bowman, B.H.; Yang, F.
Gene 49, 167-175, 1986
A;Title: The human transferrin gene: 5' region contains conserved sequences which match
A;Reference number: A29090; MUID:87192006; PMID:3106157
A;Accession: A29090
A;Molecule type: DNA
A;Residues: 1-72;291-300 <ADR>
A;Cross-references: EMBL:MI5673
R;Uzan, G.; Frain, M.; Park, I.; Besmond, C.; Maessen, G.; Trepot, J.S.; Zakin, M.M.; Ka
Biochem. Biophys. Res. Commun. 119, 273-281, 1984
A;Title: Molecular cloning and sequence analysis of cDNA for human transferrin.
A;Reference number: A32739; MUID:84153910; PMID:6322780
A;Accession: A32739
A;Molecule type: mRNA
A;Residues: 422-690 'G', 692-698 <UZA>
A;Cross-references: EMBL:MI2525; NID:g339468; PIDN:AAA61142.1; PID:g339469
R;MacGillivray, R.T.A.; Mendez, E.; Sinha, S.K.; Sutton, M.R.; Linsback-Zins, J.; Brew,
Proc. Natl. Acad. Sci. U.S.A. 79, 2504-2508, 1982
A;Title: The complete amino acid sequence of human serum transferrin.
A;Reference number: A93911; MUID:92222166; PMID:6953407
A;Contents: annotation; disulfide bonds
R;Hersberger, C.L.; Larson, J.L.; Arnold, B.; Rosteck, P.R.
Ann. N. Y. Acad. Sci. 646, 140-154, 1991
A;Title: A cloned gene for human transferrin.
A;Reference number: I51959; MUID:92231399; PMID:1809186
A;Accession: I51959
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-698 <RES>
R;Duguid, J.R.; Bohm, C.W.; Liu, N.G.; Tourtellotte, W.W.
Proc. Natl. Acad. Sci. U.S.A. 86, 7260-7264, 1989
A;Title: Changes in brain gene expression shared by scrapie and Alzheimer disease.
A;Reference number: I48174; MUID:89386721; PMID:2780570
A;Accession: I63133
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 636-696 <RES>
A;Cross-references: GB:M26641; NID:g339988; PIDN:AAA61233.1; PID:g339989
R;Schaeffer, E.; Lucero, M.A.; Jeltsch, J.M.; Py, M.C.; Levin, M.J.; Chambon, P.; Cohen,
Gene 56, 109-116, 1987
A;Title: Complete structure of the human transferrin gene. Comparison with analogous chi
A;Reference number: I54011; MUID:88056305; PMID:3678832
A;Accession: I54011
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-72 <RES>
A;Cross-references: GB:M17611; NID:g339480; PIDN:AAA61147.1; PID:g339485
A;Accession: I68160
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 564-698 <REA>
A;Cross-references: GB:M17614; NID:g339483; PIDN:AAA61148.1; PID:g339486
C;Comment: Each of the two repetitive domains binds a ferric ion and a bicarbonate anion
C;Genetics:
A;Gene: GDB:TF
A;Cross-references: GDB:120432; OMIM:190000
A;Map position: 3q21-3q21
A;Intons: 15/1; 72/3; 119/1; 168/1; 212/2; 231/1; 290/3; 350/1; 401/3; 433/1; 444/1; 49
C;Function:
A;Description: binds iron for delivery into cells
C;Superfamily: transferrin; transferrin repeat homology
C;Keywords: duplication; glycoprotein; iron transport; metal binding; plasma
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-698/Product: transferrin #status experimental <MAT>

F;20-350/Domain: transferrin repeat homology <TRH1>
F;356-686/Domain: transferrin repeat homology <TRH2>
F;58-67,38-58,137-213,156-350,177-193,180-196,190-198,246-260,358-615,364-396,374-387,4
F;432,630/Binding site: carbohydrate (Asn) (covalent) #status experimental
Query Match 32.3%; Score 1269.5; DB 1; Length 698;
Best Local Similarity 40.4%; Pred. No. 1.4e-88;
Matches 304; Conservative 118; Mismatches 221; Indels 109; Gaps 32;
QY 1 MRGPGALML--LLALRTVLGMEVWCATSDPEQHKCGNMSEAFRE-----AGIOPSL 52
DB 1 MRLAVGALLVCAVLGLCLAVDPKTVKCAVSEHATKC-----OSFRDHMKSVTPSDGPSV 56
QY 53 LCVRGTSADHCQVLIQAQADAITLDGGAIYEAG-KEHGLKPVVGEVY--DQEVGTSYYA 109
DB 57 ACVKKASYLDCIRAIANAEDAVTLDAGLVYDAYLAPNNLKPVVAEFYSGSKEDPQTFYYA 116
QY 110 VAVRRSSHVTIDTLKGVSCHTGINRTVGNVVPVGLVSGRLSVMGCD-----VLK 162
DB 117 VAVVKXDSGFQNMQLRGKSKCHTGLGRSAGWNIPIGLLY-----CDLPFRKPLEK 167
QY 163 AVSDYFGSCVPCGAGTSYSESLCRICRGDSSGEGVCDKSLERYDYDYGAFRCLEAGAG 222
DB 168 AVANFFSGCAPCADGTDFFQ-LCQLCPG-----CGCSTLQYFGYSGAFKCLKDGAG 219
QY 223 DVAFVXSTVLENTDGTLPFSGQALLSOD-FELLCRDGSRADVTWROCHLARVPAHAV 281
DB 220 DVAFVXSTIFENLANKA-----DRDQYELLCNDTRKPVDEYKDCCHLAQVPSHTV 270
QY 282 VVRADTDGG---LIFRLNEGQRLFSHEGS-SFQMFSSSEAYGOKDLFKDSTSELVPIAT 337
DB 271 VAR--SMGKEDLIWELNQAOEHFGKDKSKEQLFSS-PHG-KDLLFKDSAHGFLKVPV 326
QY 338 Q-TYEAWLGHVYLHMKGLL-----CDPNRLPPVLRMCVLSLTPRIQKCGDMAVA 385
DB 327 RMDAKMYLGYEYTAIRNLRGTCPPEAPTDECKP-----VKWALSHHRLKCDMSV- 379
QY 386 FRQRILKPIQCVSAKSPQHCMERIOAEQVDAVTLTSGEDIYTAGKYGVLVPAAGEHYAPE 445
DB 380 ----NSVGKIECVSAETTEDCIKINNGEADAMSLOGGFYIAG-KGLVLPVLAENYKS 434
QY 446 DSN-----SYVVAVVRDSSHAFTLDELGRKRSCHAGFGSPAGWDVPVGLIQRGPIR 500
DB 435 DNCEDTPEAGYFAVAVVKKASD-LTWDNLKGGKSKCHTAVGRTAGWNIPMGLL-----YNK 489
QY 501 PKDCDVLTVASFEFNASCVPVNNPNYPSLLCALCVGDEGRNKCVCNSOERYVYRGAF 560
DB 490 INHC-----RFDEFFSEGCAPGSKK-----DSSLCKLCMG--SGLNLCBPNKKEGYGTGAF 540
QY 561 RCLIVENAGDVAFVRHRTVFDNTNGHNSPEWAAELRSEDEYELLCPCNGARAEVSQFAACNLA 620
DB 541 RCLVEK-GDVAFVKHQTVPQNTGGKPNPFWAKNLNEKYELLCCLDCTRKPVVEYANCHLA 599
QY 621 QTPPHAVWRPDTNIPVYVGLDKAQDLFGDDHN--KNGFKMPDSSNYHGQDLFLKDATV 678
DB 600 RAPNHAVVRKDKEA-CVHKILLRQOHLFGSNVTDGSGNFCFLRSET---KDLLFRDDTV 655
QY 679 RAVPVCEKTYRGWGLDYVAALGEGSSQOCS 710
DB 656 CLAKLHDRNTYKYLGEYVKAIGNL--RKCS 685

RESULT 4

S33761
transferrin precursor - horse
N;Alternate names: growth-promoting factor
C;Species: Equus caballus (domestic horse)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
C;Accession: S33761; S02145
R;Carpenter, M.A.; Broad, T.E.
Biochim. Biophys. Acta 1173, 230-232, 1993
A;Title: The cDNA sequence of horse transferrin.
A;Reference number: S33761; MUID:93277958; PMID:8504171

A:Accession: S33761
A:Molecule type: mRNA
A:Residues: 1-706 <CAR>
A:Cross-references: EMBL:M69020; NID:G164242; PIDN:AAA30958.1; PID:G164243
A:Experimental source: liver; developmental stage adult
R:Yoshinari, K.; Yuasa, K.; Iga, F.; Mimura, A.
Biochim. Biophys. Acta 1010, 28-34, 1989
A:Title: A growth-promoting factor for human myeloid leukemia cells from horse serum ide
A:Reference number: S02145; MUID:89076897; PMID:2909248
A:Accession: S02145
A:Molecule type: protein
A:Residues: 20-35, 'X', '37', 'X', '39-40', 'X', '43-44' <YOS>
C:Complex: monomer
C:Superfamily: transferrin; transferrin repeat homology
C:Keywords: duplication; glycoprotein; iron transport; metal binding; plasma
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-706/Product: transferrin #status experimental <MAT>
F:358-694/Domain: transferrin repeat homology <TRH2>
F:26-64, 36-55, 134-215, 174-190, 177-198, 187-200, 248-262, 360-623, 366-398, 376-389, 423-701, 44

Query Match 32.1%; Score 1263.5; DB 2; Length 706;
Best Local Similarity 41.2%; Pred. No. 4.1e-88;
Matches 307; Conservative 113; Mismatches 237; Indels 89; Gaps 30;

QY 1 MRGPGALWLLALRTVLGGMVRVCATSDPEQKCGMSEAFREAGIOPSL-CVRGTS 59
DB 1 MLALRALACAVLGLCLAEQTVCTVSNHEVSKASPRDSMKSIIVAPPLVACVKRTS 60
QY 60 ADHCVQLIAQAQADAITLDGGAIEAG-KEHGLKPVGVGVYDQ--EVGTSYVAVAVRRS 116
DB 61 YLECIKAITADNEADAVTLDAVLFEAGLSPLYNKPVAEFYGSKTEPQTHYVAVAVKKN 120
QY 117 SHVTIDTLKGVSKCHTGNRTVGNVVPVGYL---VESGRLSVMGCDVLKAVSDYFGGSCV 173
DB 121 SNFNQLQKQKSKCHTGLGRSAGNIPIGLLYQWLPPEPSLIQ-----KAVSNFFAGSCV 175
QY 174 PGAGTYSSESICRLCRGDSGEGVCDKSPLERYDYSGAPCLAEAGDGVAFVKGSTVL 233
DB 176 PCADRTAV-PNVCQLCVG--KGTDKACSNHEPYFGYSGAFKCLADGAGDGVAFVKGSTVL 232
QY 234 ENTGDKTLPSWGQALLSDQFELLCRDGRADYVTEWRQCHLARPVAHVAVVRA-DTDDGGLI 292
DB 233 EN-----LP--QADRDVEYQLLDRNTRKSDVEYKDCYLASIPSHAVVARSVDGKEDI 284
QY 293 FLLNEGQRLFSHESG-SQFMSSSAYGQKOLLFKDSTSEL---VPIATQTYEAWLGHVYL 349
DB 285 WGLLNAQAQSHFGTEKSKDPLFESS-PHG-KOLLFKDSALGFLRIPPAMDITW-LYLGYEYV 341
QY 350 HAMKGLLCD-----NRLPPYLRCVLSPTPEIKQCDMAVAFRRQRLKPEIOCVSAKSPQ 404
DB 342 TAIURLREDIRPEVPDECKVKVCAIGHHEKVKCDEWSV-----NSGGNIECESAQSTE 396
QY 405 HCWERTIQAQVDVATLSGSDIYTAGKYGLVPAAGEHYAPEDSS-----NSYVYVAVV 457
DB 397 DCIAIKVGEADAMSIDGFFIYIAG-KCGLVPVLAENYETRSGSACVDTPEGYHAVAV 455
QY 458 RRDSSHAFTLDELGRKRSCHAGFGSPAGWDVPVGVGALIORGFTRPKDCDVLTVAVSEFFNAS 517
DB 456 KSSSDPDLTWSNLKGKSKCHTGVDRTAGNIPMGLL-----YSEIKHCE-----FDKFFREG 507
QY 518 CVPVNNPKNYPSSLCAICVGDQGRNK-CVGNQSERYCYRCFAFCLVENAGDVAFVRHT 576
DB 508 CAP-GYRRN--STLNCILGSGASGPGRECEPNHHERYGYTGAFCFLVEK-GDVAFVRHQ 563
QY 577 TVFDNTNGHNSFPWAAELRSEDIYELCPNGARAEVYSQFAACNLAIQIPHAVMVRDPTNIF 636
DB 564 TVEQNTDGRNPDMAKWLKSENFKLLCPDGTGRKSTFKSCYLARAPNHAHVSRK-----618
QY 637 TVYGLLDKQAQDLFGDDHN-----KNG-----FKMFSSNVHGGQDLFKDIAVRAVPVG 684
DB 619 -----EKAACVCQELHNGQASYGKNGSCHCPDKFLFQSAT---KDLLFRDDTQCLANLQ 669
QY 685 EKTYYRGMGLDYVAALGCMSSQOCS 710

DB 670 PTTTYKTYLGEKYLTAVALN--RQCS 693
||||: ||| | : | : | : |||
RESULT 5
TPHUL
lactotransferrin precursor [validated] - human
N:Alternate names: lactoferrin
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1992 #sequence revision 21-Nov-1997 #text change 08-Dec-2000
C:Accession: G01394; S11228; A45401; S10324; S15853; S20841; S07160; A61169; A31000; S74
R:Cho, Y.
submitted to the EMBL Data Library, March 1994
A:Reference number: G06820
A:Accession: G01394
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-711 <CHO>
A:Cross-references: EMBL:U07643; NID:G467236; PIDN:AAB60324.1; PID:G467237
Nucleic Acids Res. 18, 5288, 1990
R:Rey, M.W.; Woloshuk, S.L.; deBoer, H.A.; Pieper, F.R.
A:Title: Complete nucleotide sequence of human mammary gland lactoferrin.
A:Reference number: S11228; MUID:90384839; PMID:2402455
A:Accession: S11228
A:Molecule type: mRNA
A:Residues: 1-148, 'T', '150-422, 'C', '424-711' <REY>
A:Cross-references: EMBL:X53961; NID:G34415; PIDN:CAA37914.1; PID:G34416
R:Teng, C.T.; Liu, Y.; Yang, N.; Walmer, D.; Panella, T.
Mol. Endocrinol. 6, 1969-1981, 1992
A:Title: Differential molecular mechanism of the estrogen action that regulates lactoferrin
A:Reference number: A45401; MUID:93125571; PMID:1480183
A:Accession: A45401
A:Molecule type: DNA
A:Residues: 1-15 <TEN>
A:Cross-references: GB:S52659; NID:G263311; PIDN:AAB24877.1; PID:G263312
A:Experimental source: Placenta
A>Note: sequence extracted from NCBI backbone (NCBI:122202)
R:Powell, M.J.; Ogden, J.R.
Nucleic Acids Res. 18, 4013, 1990
A:Title: Nucleotide sequence of human lactoferrin cDNA.
A:Reference number: S10324; MUID:90326549; PMID:2374734
A:Accession: S10324
A:Molecule type: mRNA
A:Residues: 3-711 <POW>
A:Cross-references: EMBL:X52941; NID:G34411; PIDN:CAA37116.1; PID:G34412
R:Stowell, K.M.; Rado, T.A.; Funk, W.D.; Tweedie, J.W.
Biochem. J. 276, 349-355, 1991
A:Title: Expression of cloned human lactoferrin in baby-hamster kidney cells.
A:Reference number: S15853; MUID:91264786; PMID:2049066
A:Accession: S15853
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 20-31 <ST1>
A:Accession: S20841
A:Molecule type: protein
A:Residues: 20-28, 'X', '30-31' <ST2>
R:Rado, T.A.; Wei, X.; Benz Jr., E.J.
Blood 70, 989-993, 1987
A:Title: Isolation of lactoferrin cDNA from a human myeloid library and expression of mr
A:Reference number: S07160; MUID:88001031; PMID:3477300
A:Accession: S07160
A:Molecule type: mRNA
A:Residues: 436-487, 'A', '489-711' <RAD>
A:Cross-references: EMBL:M18642; NID:G186815; PIDN:AAA86665.1; PID:G386855
R:Panella, T.J.; Liu, Y.; Huang, A.T.; Teng, C.T.
Cancer Res. 51, 3037-3043, 1991
A:Title: Polymorphism and altered methylation of the lactoferrin gene in normal leukocytes
A:Reference number: A61169; MUID:91235214; PMID:1674448
A:Accession: A61169
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 3-701, 'SWKPVN' <PAN>
A:Experimental source: normal breast tissue

R;Metz-Boutique, M.H.; Jolles, J.; Mazurier, J.; Schoentgen, F.; Legrand, D.; Spik, G.; Eur. J. Biochem. 145, 659-666, 1984
A;Title: Human lactoferrin: amino acid sequence and structural comparisons with other
A;Reference number: A31000; MUID:85076667; PMID:6510420
A;Accession: A31000
A;Molecule type: protein
A;Residues: 20-140,142-169;171-203,'L',205,'K',207-208,'K',210-385,'Q',387-391,'W',393-4
A;Note: this is the final paper in a series
R;Houen, G.; Hoegdall, E.V.; Barkholt, V.; Norbkov, L.
Eur. J. Biochem. 241, 303-308, 1996
A;Title: Lactoferrin: similarity to diamine oxidase and purification by aminohexyl affinity
A;Reference number: S74119; MUID:97054624; PMID:898921
A;Accession: S74119
A;Molecule type: protein
A;Residues: 'G',23-24,'R',26-27,'XX',30-32 <HOU>
A;Experimental source: neutrophil granulocytes
C;Genetics:
A;Gene: GDB:LTF
A;Cross-references: GDB:119368; OMIM:150210
A;Map position: 3q21-3q23
C;Superfamily: transferrin; transferrin repeat homology
C;Keywords: duplication; glycoprotein; iron binding; milk
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-711/Product: lactoferrin #status experimental <MAT>
F;21-356/Domain: transferrin repeat homology <TRH1>
F;360-699/Domain: transferrin repeat homology <TRH2>
F;29-65,39-56,135-218,177-193,190-201,251-265,503-697,595-609/Disulfide bonds: #status
F;157,498/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;368-400,378-391,425-706,447-669,479-554,513-527,524-537,647-652/Disulfide bonds: #stat

Query Match 32.0%; Score 1260.5; DB 1; Length 711;
Best Local Similarity 39.8%; Pred. No. 7.1e-88;
Matches 302; Conservative 115; Mismatches 234; Indels 107; Gaps 29;

QY 6 GALMLLLAURTLLGGMEVWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVQ 65
DB 12 GAGLCLAGRR-----RSVOMCAVQSEATKCFQWQRMKRV-PPVSCIKRDSPTQCIQ 67
QY 66 LIAAQEADAITLDGGAIEAG-KEHGLKPVVGVVY--DOEVGTSYAVAVVRRSSHVTID 122
DB 68 AIAENRADAVTLGGFIYEAGLAPYKLRPVAAEVYGTQPRHYAVAVVKKGSFQLN 127
QY 123 TLKGVKSCHTGINTRTVGMNVPVGYL-----VESGRLSVMGCDVLKAVSDYFGG 170
DB 128 ELQGLKSCHTGLRTAGNVPVIGTLRFLNWTGPPPIEA-----AVARFFSA 175
QY 171 SCVPGAGETSYSLCRLCRGDSGEGVCDKSPLEYYDYGAFRCLEAGAGDAVFKHS 230
DB 176 SCVPGADKQGF-PNLCLRCAG--TGENKCAFSSQBPYFSYGAFKCLRDGAGDAVFIRES 232
QY 231 TVLENTDGKTLPSWGOALLSQDFELLCHRDGSRADVTETWROCHLARVPAHVVVRA-DTDG 289
DB 233 TVPFDLSDEA-----ERDEVELLCPNTRKPVDFKDXCHLARVPSHAVVARSVNGKE 284
QY 290 GLIFRLNIGQRLFSHEGS-SFQMFSSAYGOKDLFLKOST---SELVP-IATQTYEAWL 344
DB 285 DAIWNLLRQAEKFGKDKSPKOLFQSPS-GQKDLFLKOSTSAIGFSRVPRIDSGLY--L 340
QY 345 GHEYLHMKGLLCPNRLP---PYLRWCVLSTPEIKQCGDMAVAFRRQRLKPEIQCVSAK 401
DB 341 GSGYFTAIQNLKSEEEVARRARVVMCAVGEOLKCNOWS-----GLSEGSVTCSSAS 395
QY 402 SPOCHMERIOAEQVDVTLTSGEDIYTAGKKYGLVPAAGEHYAPEDSSN-----S 450
DB 396 TTEDCIALLVKGADAMSLDGGVYTAG-KCGLVPVLAENYKSKQSSDDPNPCVDRPVBG 454
QY 451 YVAVVRRDSSHAFTLDELGRKRSCHAGFGSPAGWDVPVVGALIQRFIRPKDCDVLTA 510
DB 455 YLAVAVVRR-SDTSLTWSNVKSKSCTAVDRTAGWNIEMGLL----FMQTGSC----KF 505
QY 511 SEFFNASCVPVNNPKYPSLCLALCVGDEGRNKCVCNQSERYGYRGAFRCLEVENAGDV 570
DB 506 DEYFSQSCAPGSDPR---SNLCALCTGDEGENKCVNPSNERYGYGTGAFRCLEVENAGDV 562

QY 571 AFVRHTTVPDNTNGHNSPEWAAELRSEDYELLCPNGARAEVSQFAACNLAIQIPPHAVMR 630
DB 563 AFVKDVTVLQNTDGNNEAWAKDLADFALLCLDKRKPVTREARSCHLAWAPNHAVVS 622
QY 631 PDTNFTVYVGLDKKQADLFGDDHKNKG-----FKMFDSSNYHGQDILLFKDQATVRAVPVG 684
DB 623 MD-KVERLKVLLHQQAQFG-----RNGSDCPDKFCLFQSET---KNLLFNDNTECLARLH 674
QY 685 EKTITRGWGLDVAALEGMSQQSCSGAAAPAPGAPLL 722
DB 675 GKITYEKYLGPOYVAGITNL--KKCS-----TSPLL 703

RESULT 6
JC2323
lactoferrin - goat
C;Species: Capra aegagrus hircus (domestic goat)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 07-May-1999
C;Accession: JC2323
R;Le Provost, F.; Nocart, M.; Guerin, G.; Martin, P.
Biochem. Biophys. Res. Commun. 203, 1324-1332, 1994
A;Title: Characterization of the goat lactoferrin cDNA: assignment of the relevant loci
A;Reference number: JC2323; MUID:94380047; PMID:8093048
A;Accession: JC2323
A;Molecule type: mRNA
A;Residues: 1-708 cLEP
C;Superfamily: transferrin; transferrin repeat homology
C;Keywords: duplication; glycoprotein
F;359-696/Domain: transferrin repeat homology <TRH2>
F;252,300,387,495,564/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 32.0%; Score 1258; DB 2; Length 708;
Best Local Similarity 40.8%; Pred. No. 1.1e-87;
Matches 305; Conservative 110; Mismatches 243; Indels 90; Gaps 30;

QY 6 GALMLLLAURTLLGGMEVWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVQ 65
DB 12 GAGLCLAGRR-----KNVWCAISLPWMSKCYQWQRMKLG-APSICTVRRTSALECIR 66
QY 66 LIAAQEADAITLDGGAIEAGKE-HCLKPVVGEVYDOEVG--TSYVAVAVVRRSSHVTID 122
DB 67 AIAENRADAVTLDSGMVFEAGDPYKLRPVAAEIVGTETKSPQTHYAVAVVKKGSNPKLD 126
QY 123 TLKGVKSCHTGINTRTVGMNVPVGYL-----VESGRLSVMGCDVLKAVSDYFGGCVPA 176
DB 127 QLOQKSCSHMGLGRAGWNI PVGILRPPLSWTESAE-----PLQAVARFFSASCVPVC 180
QY 177 GETSYSESICRLCRGDSGEGVCDKSPLEYYDYGAFRCLEAGAGDAVFKHSTVLENT 236
DB 181 DGKAY-PNLCLQCKG--VGENKCACSSQBPYFGYSGAFKCLQDQAGDAVFKETTVFEN- 236
QY 237 DKTLPSPWGOALLSQDFELLCHRDGSRADVTETWROCHLARVPAHVVVRA-DTDGGLIFRL 295
DB 237 ----LPEKAD---RDQYELLCLNTRAPVDAFKECHLAQVPSHAVVARSVDGKENLIWEL 289
QY 296 LNEGQRLF-SHEGSSQMFSSAYGOKDLFLKOSTSELVPIATQTYEA-WLGEHYLHMK 353
DB 290 LRKAQEKFKNKSQRFQFSGPE-GRDILLFKDSALGFVRIPSKVDLSALYLSRYLTALK 348
QY 354 GLLCDPNRLP---PYLRWCVLSTPEIKQCGDMAVAFRRQRLKPEIQCVSAKSPQHMERI 410
DB 349 NLRTAEELKARCTRVWCAVGPPEQSKQOWS-----EQSGONVTCATASTTDDCIALV 403
QY 411 QAEQVDVTLTSGEDIYTAGKKYGLVPAAGEH-----YAPED-----SSNSYVAVVRRDS 461
DB 404 LKGEADALSLDGGYITAG-KCGLVPVMAENKSKYSLSLDCVLRPTEGYLAVAVVKK-A 461
QY 462 SHAFTLDELGRKRSCHAGFGSPAGWDVPVVGALI-ORGFIRPKDCDVLTVSEFFNASCVP 520
DB 462 NEGLTWSNLKSKSCTAVDRTAGWNIEMGLTANQTG-----SC-----AFDEFFSOSCAP 512
QY 521 VNNPKYPSLCLALCVGDEGRNKCVCNQSERYGYRGAFRCLEVENAGDAVFRHTTTFD 580

Db 513 GADPK---SSLCALCAGDQGLDKCPNSKERYGYGTGAFCCLAEDVGVAFVKNDTWWE 569
QY 581 NTNGHNSPWAELRSYDYLCPNGARAEVSQFACNLAQIPPHAVMVRPDTNIFTVYG 640
Db 570 NTNGESSADWAKNLNREDFRLCLDGTTPVTEAOSCYLAVAPNHAIVNSRSDRAAHVEQV 629
QY 641 LLDKQDLRGDDHNKG-----EKMFDSNNYHGQDLFPKDATVRAVPVGKTYRGWLG 694
Db 630 LHQO-QALFG-----KNGKVCPOQCLFKSET---KNLFDNDTECLAKLGGRTYKYL 681
QY 695 LDYVAALQEGSSQCSGGAAPAGPALL 722
Db 682 TEYVTAIANL--KKCS-----TSPLL 700
RESULT 7
TFEOL
lactotransferrin precursor - bovine
N:Alternate names: lactoferrin
C:Species: Bos primigenius taurus (cattle)
C:Date: 31-Mar-1992 #sequence revision 21-Nov-1997 #text_change 11-May-2000
A:Accession: I45919; S14674; S14110; S18517; JT0595; S13097; S18518; S13881; PL0148; S21
R:Tsang, T.C.; Burns, D.K.; Wang, F.; Pan, Y.
FASEB J. 6, 233, 1991
A:Title: Cloning of a 80-kD advanced glycosylation end product (AGE) binding protein fr
A:Reference number: I45919
A:Accession: I45919
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-708 <TSA>
A:CROSS-references: GB:L08604; NID:gl63269; PIDN:AAA30609.1; PID:gl63270
R:Pieper, A.
submitted to the EMBL Data Library, November 1990
A:Reference number: S14674
A:Accession: S14674
A:Molecule type: mRNA
A:Residues: 1-144, 'V', 146-163, 'PP', 166-339, 'A', 341-438, 'Y', 440-513, 'R', 515-708 <PI1>
A:CROSS-references: EMBL:X57084; NID:G505; PIDN:CAA40366.1; PID:G506
R:Pieper, A.; Colavizza, D.; Benaisa, M.; Maes, P.; Tartar, A.; Montreuil, J.; Spik, G.
Eur. J. Biochem. 196, 177-184, 1991
A:Title: Molecular cloning and sequence analysis of bovine lactotransferrin.
A:Reference number: S14110; MUID:91160550; PMID:2001696
A:Accession: S14110
A:Molecule type: mRNA
A:Residues: 3-144, 'V', 146-339, 'A', 341-438, 'Y', 440-513, 'R', 515-708 <PI2>
A:CROSS-references: EMBL:X57084
A:Accession: S18517
A:Molecule type: protein
A:Residues: 20-35, 'DS', 114-148-163, 'PP', 166-178, 'V', 'P', 183-190, 205-212, 230-239, 304-339, 59
R:Goodman, R.E.; Schanbacher, F.L.
Biochem. Biophys. Res. Commun. 180, 75-84, 1991
A:Title: Bovine lactoferrin mRNA: sequence, analysis, and expression in the mammary gland
A:Reference number: JT0595, MUID:92028986; PMID:1718281
A:Accession: JT0595
A:Molecule type: mRNA
A:Residues: 1-65, 'FG', 68-296, 'S', 298-339, 'A', 341-708 <GOO>
A:CROSS-references: GB:IM63502
A:Note: the authors translated the codon CCG for residue 66 as Arg and TCT for residue 2
R:Mead, P.E.; Tweedie, J.W.
Nucleic Acids Res. 18, 7167, 1990
A:Title: cDNA and protein sequence of bovine lactoferrin.
A:Reference number: S13097; MUID:91088328; PMID:2263492
A:Accession: S13097
A:Molecule type: mRNA
A:Residues: 28-33, 'DS', 36-38, 'P', 40-708 <MEA>
A:CROSS-references: EMBL:X54801
A:Accession: S18518
A:Molecule type: protein
A:Residues: 20-47, 59-66, 132-139, 256-277, 278, 305-332, 343-351, 361-363, 586, 587-589, 598-619
R:Mead, P.E.
submitted to the EMBL Data Library, October 1990
A:Reference number: S13881

A:Accession: S13881
A:Molecule type: mRNA
A:Residues: 28-38, 'P', 40-86, 'C', 88-708 <ME3>
A:CROSS-references: EMBL:X54801
R:Rejman, J.J.; Hegarty, H.M.; Hurley, W.L.
Comp. Biochem. Physiol. B 93, 929-934, 1989
A:Title: Purification and characterization of bovine lactoferrin from secretions of the
A:Reference number: PL0148; MUID:90031466; PMID:2805645
A:Accession: PL0148
A:Molecule type: protein
A:Residues: 20-27, 'X', 29-37, 'X', 39-54, 'X', 56-59 <REU>
R:Beilamy, W.; Takase, M.; Yamauchi, K.; Wakabayashi, H.; Kawase, K.; Tomita, M.
Biochim. Biophys. Acta 1121, 130-136, 1992
A:Title: Identification of the bactericidal domain of lactoferrin.
A:Reference number: S21756; MUID:92287941; PMID:1599934
A:Accession: S21756
A:Molecule type: protein
A:Residues: 36-60 <BEL>
R:Shimazaki, K.; Tanaka, T.; Kon, H.; Oota, K.; Kawaguchi, A.; Maki, Y.; Sato, T.
J. Dairy Sci. 76, 946-955, 1993
A:Title: Separation and characterization of the C-terminal half molecule of bovine lacto
A:Reference number: A56659; MUID:93253156; PMID:8486845
A:Accession: A56659
A:Molecule type: protein
A:Residues: 20-25, 302-308, 359-366, 'X', 368-376, 'X', 378 <SHI>
C:Superfamily: transferrin; transferrin repeat homology
C:Keywords: duplication; glycoprotein; iron; iron binding; metalloprotein; milk
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-708/Product: lactotransferrin #status experimental <MAT>
F:20-355/Domain: transferrin repeat homology <TRH1>
F:36-60/Region: antimicrobial
F:359-696/Domain: transferrin repeat homology <TRH2>
F:28-64, 134-217, 176-192, 179-200, 189-202, 250-264, 367-399, 377-390, 424-703, 444-666, 476-551,
F:78-55/Disulfide bonds: #status predicted
F:79, 111, 211, 272/Binding site: iron (Asp, Tyr, Tyr, His) #status experimental
F:140/Binding site: carbonate (Arg) #status experimental
F:252, 300, 387, 495, 564/Binding site: carbohydrate (Asn)
F:414, 452, 545, 614/Binding site: iron (Asp, Tyr, Tyr, His) #status experimental
F:482/Binding site: carbonate (Arg) #status experimental
Query Match 31.7%; Score 1249; DB 1; Length 708;
Best Local Similarity 40.2%; Pred. No. 5.3e-87;
Matches 300; Conservative 119; Mismatches 241; Indels 86; Gaps 30;
QY 6 GALMILLALRTVLGMEVRCWATSDPEOHKCGNNSAPREAGIOPSLCLVCGTSDHCVQ 65
Db 12 GALGLCLAAPR-----KNVRWCTISQPEWFKRRWRMKLG-AP SITCVRRFALECI 66
QY 66 LIAAQEADAITLDGGAIEYACKE-HGLKPVVGEVY--DOEVGTSTYAVAVVRRSSHVTID 122
Db 67 AIAEKKADAVTLGDMVFEGARDPYKURPVAEYIGTKESPTQTHYAVAVVKKGSNFQID 126
QY 123 TLKGVKSCHTGINRTVGNVNPVGLVSGRLSVMGCDVLK----AVSDYFGSGCVPFGAGE 178
Db 127 QLQGRKSCHTGLGRSAGWIIPMGIL----RPLVSWTESLEPLQGAVERFSSASCPVCIDR 182
QY 179 TSYSESICRLCRGDSSSGEGVCDKSPLERYDYSGAFRCIAGAGDVAFVKHSTVLENTDG 238
Db 183 QAY-PNLQCLCKGE--GENQCACSRSEPFYFGYAFKCLQDGAGDVAFVKETTVFEN--- 236
QY 239 KTLPSWGQALLSQDFELLCRDGSRADVTVEHQCHLARVPAHAVVVR--DTDGGGLIFRLN 297
Db 237 --LPEKAD---RDQYELLCNLNNSRAPVDAFKCHLAQVPSHAVVARSVDGKEDLWLKLS 291
QY 298 EQGRLF-SHEGSSQFMFSSEAYGQKDLIFKDSITSELVDIATQTYEA-WLGEHYLHAMKGL 355
Db 292 KAEKFGKNKRSQQLFGSP-PGQDILLFKDSALGLFLRIPSKVDLSALYLSRYLTTLKNL 350
QY 356 LCDNRLRPP-VLR--WCVLSTPTETQCGDMAVAFRRQRLKPEIQCVSAKSPQHMERQA 412
Db 351 RETAEVKARYTRVVMCAVGPGEQKCCQWS-----QOSQONVTCATASTTDDCIVLVLK 405
QY 413 EQVDVLTSGEDIYTAGKKYGLVPAAGEHYAPEDSSN-----SYVVAVVVRDSSH 463

Db 406 GEADALNDGGYIYTAG-KGGLVPLVLAENRKSXKSHSLDCVLRPTGTYLAVAVVKK-ANE 463
 QY 464 AFTLDLRLGRKSRCHAGFGSPAGDVPVVGALI-ORGFIKPKDCDVLTAVSEFFFNASCVPVN 522
 Db 464 GLTWSLAKOKKSCHTAVDRTAGWNI PMGLIVNTG-----SC-----AFDEFFSQSCAPGA 514
 QY 523 NPKNYPSSLCALCVGDEQGRNKCVCNSQERYGYRGAFCRLVFNAGDVAFVRHTTVFDNT 582
 Db 515 DPR--SRLCALCAGDDGLDKCVPNSKEKYGYTGAFRCLEADVDGVAFKNDTWENT 571
 QY 583 NGHNSPWAELRSEDEYELLCPNARAEVYQFAACNLAQIIPPHAVMVRPDNTIFTYVGLL 642
 Db 572 NGESTADWAKNLRNEDFRLLCLDGRKPVTEAQSCHLAVAPNHAVVRSRDRAHVKQVLL 631
 QY 643 DKAQDLFGDDHNKNG-----FKMFDSSNYHGQDLLFKDQATVRAVPVGEKTTVYRGMLGLD 696
 Db 632 HQ-QALFG---KNGKNCPCFKCLFKSET---KNLLFNNDTECLAKLGGRPYEEYLGTE 683
 QY 697 YVAALGSMSSQCGSAAAPAGAPLL 722
 Db 684 YVTAIANL--KKCS-----TSPLL 700

RESULT 8
 TFRBP

transferrin precursor - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text change 22-Jun-1999
 C:Accession: S16246; A61239; C61573; S00335; S02694; A26504; S14853
 R:Banfield, D.K.; Chow, B.K.C.; Funk, W.D.; Robertson, K.A.; Umelas, T.M.; Woodworth, R.
 Biochem. Biophys. Acta 1089, 262-265, 1991
 A:Title: The nucleotide sequence of rabbit liver transferrin cDNA.
 A:Reference number: S16246; UID:91274362; PMID:2054387
 A:Accession: S16246
 A:Molecule type: mRNA
 A:Residues: 1-694 <BAN>
 A:Cross-references: EMBL:X58533; NID:g1750; PIDN:CAA41424.1; PID:g1751
 R:Pierpaoli, W.; Dall'Ara, A.; Yi, C.; Neri, P.; Santucci, A.; Choay, J.
 Cell. Immunol. 134, 225-234, 1991
 A:Title: Iron carrier proteins facilitate engraftment of allogeneic bone marrow and endothelial cells.
 A:Reference number: A61239; UID:91191584; PMID:2013104
 A:Accession: A61239
 A:Molecule type: protein
 A:Residues: 19-36 <PIE>
 R:Chung, M.C.M.; Chan, S.L.; Shimizu, S.
 Int. J. Biochem. 23, 609-616, 1991
 A:Title: Purification of transferrins and lactoferrin using DEAE Affi-Gel Blue.
 A:Reference number: A61573; UID:91293379; PMID:2065820
 A:Accession: C61573
 A:Molecule type: protein
 A:Residues: 19-26,'X',28-36,'X',38-53 <CHU>
 R:Godovac-Zimmermann, J.
 Biol. Chem. Hoppe-Seyler 369, 93-96, 1988
 A:Title: Isolation, characterization and N-terminal amino-acid sequence of rabbit transferrin.
 A:Reference number: S00335; UID:88209278; PMID:3365331
 A:Accession: S00335
 A:Molecule type: protein
 A:Residues: 19-45,'S',47-48,'Y',50 <GOD>
 R:Evans, R.W.; Aitken, A.; Patel, K.J.
 FEBS Lett. 238, 39-42, 1988
 A:Title: Evidence for a single glycan moiety in rabbit serum transferrin and location of the N-glycanase cleavage site.
 A:Reference number: S02694; UID:89005676; PMID:3169252
 A:Accession: S02694
 A:Molecule type: protein
 A:Residues: 482-515,'V',517-544 <EVA>
 A:Note: S16-1le was also found
 R:Heaphy, S.; Williams, J.
 Biochem. J. 205, 611-617, 1982
 A:Title: The preparation and partial characterization of N-terminal and C-terminal iron-transferrin - pig
 A:Reference number: A26504; UID:83074540; PMID:6816218
 A:Accession: A26504
 A:Molecule type: protein

A:Residues: 19-24,'N',26,'X',28-29,'S' <HEA>
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: duplication; glycoprotein; iron transport; metal binding; plasma
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-694/Product: transferrin #status experimental <MAT>
 F:19-349/Domain: transferrin repeat homology <TRH1>
 F:355-682/Domain: transferrin repeat homology <TRH2>
 F:27-66,37-57,136-212,155-349,176-192,179-195,189-197,245-259,357-611,363-395,373-386,4
 F:508/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 31-2%; Score 1228; DB 1; Length 694;
 Best Local Similarity 41.1%; Pred. No. 2.1e-85;
 Matches 303; Conservative 109; Mismatches 225; Indels 100; Gaps 34;

QY 7 ALWLLIALTLVGLGMEVRWCATSDPEQHKCGN-----MSEAFREAGIQPSLLCVRGTSADH 62
 Db 12 ALGLCLAVTE---KTVRCVANDHEASKANFRSMKKVLPEDG--PRIICVKKASYLD 65
 QY 63 CVQLTAAQADAITLDGGAIYEAG-KEHGLKPVWGEVYDQEVG--TSYVAVAVVRSSHV 119
 Db 66 CIKATAAHEADAVTLDAGLVHEAGLTNNLKPVAEFYGSKENPKTFYVAVALVKGSNF 125
 QY 120 TIDTLGVKSKCHTGNRTVGMVPGVLYVESGRSLVWGC-----VLKAVSDYEGGSC 172
 Db 126 QNLEQKKSKCHTGLRSAGWNIPIGLY-----CDLPEPRKPLEKAVASFPFGSC 176
 QY 173 VPGAGETSYSESLCRLCRDSSGEGVCDKSPLERYDYSGAFRCLEAGDVAFVRHSTV 232
 Db 177 VPCADGADFPQ-LCQLCPG-----CGCSSVQPFYGYSGAFKCLKGLGDLVAFVQETI 228
 QY 233 LENTDGKTLPSWQALLSODFELLCDRSGRADVTEWRQCHLARVPAAHVVRRA-DTDDGGL 291
 Db 229 FEN-----LPSKDE---RDQYELLCLDNRKPKVDEYEQCHLARVPAAHVVRSDGKEDL 280
 QY 292 IFRLLNEGQRLSHSGS-SFOMFSEAYCQKLLFKDSTSELVPIATQ-TYEAWLGHYL 349
 Db 281 IWELLNQAEHFQKDKSGDFQLFSS-PHG-KNLLFKDSAYGFFKPPRMDANLYLGYEV 338
 QY 350 HAMKGLI---LCDPNRLP---PYLRWCVLSTPEIQKGDMAVAFRRQRLKPEIQCVSAKSP 403
 Db 339 TAVNLRREGIC-PDPLQDECKAVKWCALSHHERLKCDEWSVTSGGL-----TECESAETP 392
 QY 404 QHMERIOAEQVDVAVTLSGEDIYTAGKYGVLPAAGEHY-----APEDSSNYSYVAV 456
 Db 393 EDCIAKIMGAEADAMSLDGGYVYIAG-QCGLVPLVAENYESTDCKAPEE---GYLSVAV 448
 QY 457 VPRDSSHAFTLDELAKRSCHAGFGSPAGMDVPVGLIQRGPIRKCDVLTAVSEFFNA 516
 Db 449 VKK-SNPDIINWNLESGKSKCHTAVDRTAGWNI PMGLL-----YNRINHC-----RFDEFFRQ 499
 QY 517 SCVPVNNPNKYPSSLCALCVGDEQGRNKCVCNSQERYGYRGAFCRLVENAGDVAFVRHT 576
 Db 500 GCAP-GSQKN--SSLCCLCIGP---SVCAPNNREGYGYTGAFCRLVEK-GDVAFVKSO 551
 QY 577 TVFDNTNGHNSPWAELRSEDEYELLCPNARAEVYQFAACNLAQIIPPHAVMVRPDNTIF 636
 Db 552 TVLQNTGGRNSPWAELRSEDEYELLCPNARAEVYQFAACNLAQIIPPHAVMVRPDNTIF 611
 QY 637 TVYGLLD---KAQDLFGDDHNKNGKPMFDSNHYGQDLLFKDQATVRAVPVGEKTTVYRGWL 693
 Db 612 VKQKLLDLQVEYNGTAVDCSSK--FCMFHSTK---KDLLFRDDTKCLVDLRGKNTYKYL 666
 QY 694 GLDYVAALGEMSSQOCS 710
 Db 667 GADYIKAVSNL--RKCS 681

RESULT 9

S01384

transferrin - pig

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 11-May-2000

C:Accession: S01384; A60520; A61573

231	ENLPDKA-----DQDYELLCKONTRRPVDDYENCYLAQVPSHAVVARSVDGKDL	281
292	IFRLNREGORLFSHEGSS-FQMFSSAYQGXDLFPKDSSTSELVPIATO-TYEAWLGHYVL	349
282	IWELLNQAQENFGKDAEOLFSS-SHG-KDLLFTDACGLRPLRVPKMDAKLYLGYEYF	339
350	HAMKGL-LCDPNRLPPVLRVLCWSTPETQKCGDMAVAFRRQRLKPEI-QCVSAKSPQHCM	407
340	AAIQHLLRRVQGTSEFQVRVWCAVQGHETKCDSHV-----LSGGILNCNSEDTHMEDCI	393
408	ERIQAEQVDAVTLTSGEDIYTAGKYGLVPAAGEHYAPEDSS-----NSGYVVA	455
394	AAI AKGEADAMSLDGGFLYTAG-KCGLVPLVAENYLSQDKERFGSKCVNTPVEGYVVA	452
456	VVRDSSHAFTLDLRLGKRCHAGFGSPAGHDVPGALI-QRGFTRPKDCDVLTAVERSEF	514
453	VVKK-SOADLTWNLSLRGKKSCHIAVGTSGAGWIIIPMGFIYNTQSGCK-----LDSEFF	502
515	NASCVPVNNPNKNYPSSLCALCVGDEQGR--NKCVGNSQERYGYRGAFRCLVENAGDVAF	572
503	SQSCAPGSDPE--SRLCALCSGSI SQGPAHTCAPNSHEGYHGFSGALRCLVEK-GDVAF	558
573	VRHTTVFDNTNGHNSPWAALRSEDYELLCPNGARAEVSOFAACNLAQI PHHAWVWRPD	632
559	VKHPVLQNTDGRNPEAWAKDLQKEDFQLLCPDGTGRKEVTEAQSCHLAAPVSHAVVRKD	618
633	TNIETVYGLDLDKAQDLFGDDHNKNGFK--MFDSSNYHGDQLLLFKDATVRAVPVGEKTTVR	690
619	KADF-VRMRLENQOELEFG----RNGFEYWMFQLFKSSTEDLLFSDDTCLANLQDKIYVO	673
691	GWGLDLYVAALLEGMSQOC	709
674	KYLGPYQLAIANY--RQC	690

U.S. Department of Justice
Federal Bureau of Investigation
Washington, D.C. 20535

RESULT 12

TFCHE

ovotransferrin precursor - chicken

N:Alternate names: conalbumin; transferrin

C:Species: Gallus gallus (chicken)

C:Date: 24-Apr-1994 #sequence revision 30-Sep-1993 #text change 22-Jun-1999

C:Accession: A26845; A91115; A32229; A91116; A40674; B61573; A90282; S02476; A

R:Jeltsch, J.M.; Hen, R.; Maroteaux, L.; Garnier, J.M.; Chambon, P.

Nucleic Acids Res. 15, 7643-7645, 1987

A:Title: Sequence of the chicken ovotransferrin gene.

A:Reference number: A26845; MUID:88015626; PMID:3658709

A:Accession: A26845

A:Molecule type: DNA

A:Residues: 1-705 <JEL>

A:Cross-references: GB:X00407; NID:G63131; PIDN:CA468468.1; PID:G295721

R:Jeltsch, J.M.; Chambon, P.

Eur. J. Biochem. 122, 291-295, 1982

A:Title: The complete nucleotide sequence of the chicken ovotransferrin mRNA.

A:Reference number: A91115; MUID:82138851; PMID:7060577

A:Accession: A91115

A:Molecule type: mRNA

A:Residues: 1-82, 'V', 84-99, 'I', 101-153, 'W', 155-238, 'LN', 241-685, 'N', 687-705 <JF>

A:Cross-references: EMBL:X02009

A:Note: the codons given for residues 132 (AAC) and 317 (UUC) are inconsistent

R:Thibodeau, S.N.; Lee, D.C.; Palmiter, R.D.

J. Biol. Chem. 253, 3771-3774, 1978

A:Title: Identical precursors for serum transferrin and egg white conalbumin.

A:Reference number: A92229; MUID:78171533; PMID:649604

A:Accession: A92229

A:Molecule type: protein

A:Residues: 1-23 <TH>

R:Williams, J.; Eillemann, T.C.; Kingston, I.B.; Wilkins, A.G.; Kuhn, K.A.

Eur. J. Biochem. 122, 297-303, 1982

A:Title: The primary structure of hen ovotransferrin.

A:Reference number: A91116; MUID:82138852; PMID:6895872

A:Accession: A91116

A:Molecule type: protein

A:Residues: 1-50; 54-82, 'Y', 84-92; 102-146; 162-168; 170-228; 241-283; 289-333; 338-3

R; Gentili, C.; Bianco, P.; Neri, M.; Malpeli, M.; Campanile, G.; Castagnola, P.; Cancedi
J. Cell Biol. 122, 703-712, 1993
A; Title: Cell proliferation, extracellular matrix mineralization, and ovotransferrin tra
A; Reference number: A40674; MUID:93328771; PMID:8393014
A; Accession: A40674
A; Molecule type: protein
A; Residues: 20-28, 'X', 30-38, 'X', 40-44 <GEN>
R; Chung, M.C.M.; Chan, S.L.; Shimizu, S.
Int. J. Biochem. 23, 609-616, 1991
A; Title: Purification of transferrins and lactoferrin using DEAE Affi-Gel Blue.
A; Reference number: A61573; MUID:91293379; PMID:2065820
A; Accession: B61573
A; Molecule type: protein
A; Residues: 20-28, 'X', 30-38, 'X', 40-43, 'S' <CHU>
R; Kingston, I.B.; Williams, J.
Biochem. J. 147, 463-472, 1975
A; Title: The amino acid sequence of a carbohydrate-containing fragment of hen ovotransfe
A; Reference number: A90282; MUID:76039467; PMID:1172663
A; Accession: A90282
A; Molecule type: protein
A; Residues: 480-582 <KIN>
R; Ellemann, T.C.; Williams, J.
Biochem. J. 116, 515-532, 1970
A; Title: The amino acid sequences of cysteic acid-containing peptides from performic aci
A; Reference number: A90246; MUID:70141846; PMID:4907959
A; Contents: annotation; disulfide bonds
R; Williams, J.; Moreton, K.
Biochem. J. 251, 849-855, 1988
A; Title: The dimerization of half-molecule fragments of transferrin.
A; Reference number: S02476; MUID:88326225; PMID:3415649
A; Accession: S02476
A; Molecule type: protein
A; Residues: 20-23; 295-302; 336-366; 674-679, 'T', 681, 'F', 683-685, 'N', 687-705 <W12>
C; Comment: Ovotransferrin (conalbumin) and transferrin have identical protein compone
rin is synthesized in the liver.
C; Comment: Ovotransferrin has a bacteriostatic function. Its concentration in avian egg
C; Comment: Plasma apotransferrin promotes the oxidation of ferrous ions, which would oth
C; Comment: In electrophoretic and genetic studies, transferrin shows strong polymorphism
C; Genes:
A; Introns: 15/1; 69/3; 106/1; 170/1; 215/2; 234/1; 290/3; 353/1; 404/3; 436/1; 448/1; 50
C; Superfamily: transferrin; transferrin repeat homology
C; Keywords: duplication; egg white; glycoprotein; iron binding; plasma
F; 1-19/Domain: signal sequence #status experimental <SIG>
F; 20-705/Product: transferrin #status experimental <MAT>
F; 21-355/Domain: transferrin repeat homology <TRH1>
F; 359-692/Domain: transferrin repeat homology <TRH2>
F; 29-64, 134-216, 178-193, 190-201, 247-261, 367-399, 424-699, 440-662, 473-549, 497-690/Disulfid
F; 39-55, 377-390, 589-603/Disulfide bonds; #status predicted
F; 492/Binding site: carbohydrate (Asn) (covalent) (partial) #status experimental
F; 507-521, 518-532/bisulfide bonds: (or 507-518, 521-532) #status predicted
Query Match 30.6%; Score 1202.5; DB 1; Length 705;
Best Local Similarity 39.4%; Pred. No. 1.8e-83;
Matches 283; Conservative 112; Mismatches 249; Indels 75; Gaps 28;
QY 23 VFWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRTSDADHCVQLIAQEAADITLDGAI 82
DB 26 IWCITISPEEKKCNLRDLTOQERI--SLTCVQKATYLDCTIKAIANNEADAISLDGGA 83
QY 83 YEAG-KEHGLKPVGVGVDOEVG--TSYAVAVVRRSHVTTDTLKVKSCHTGLNRTVG 139
DB 84 FEAGLAPYKLIKIAIEVTEHTGTSYAVAVVKKGTFTVNDLQKTSCHTGLGRAG 143
QY 140 MNVPVGLVESGRLSVMGCD---VLKAVSDYFGSCVPGAGTYSYSLCRLCRGDSSGE 196
DB 144 WNPICTLLHRAIEWEGTESGVQAVAKFFSASCPGA---TTEQKLCKRCKGDPKTK 200
QY 197 GVCDSPLERYDYSGARCLAEAGDVAFVKGSTVLENTDGTLPSPGQALLSDQFELL 256
DB 201 --CARN--APYSYSGAFHCLKDGKGDVAFVHTTNNAPDQ-----KDEYELL 246
QY 257 CRDGSRADTWEVRQCHLARVPAHVVVRADTDGGLIFRLNQRQLFSHE-GSSQFQFSS 315

DB 247 CLDGRQPDVNYKTCNWARVAHAHVARDNKNVEDIWFSLSKAOSDFGVDTKSDHFLGCP 306
QY 316 EAYGO---KDLFKDSTSEL--VPIATQTYEAWLGHYHAKMGLLCD---PNRLPPYLR 367
DB 307 PGKDPVLKDLFLKDSAIMLKRPV-SLMDSQLYLGFEYYSIAQSMRKDQLTPSPRENRIQ 365
QY 368 WCVLSTPEIQKCGDWAVAFRRQRLKPEIQCSAKSPQCMERIOAQEVDVAVTLGSEDIYT 427
DB 366 WCAVGKDSKCDRWSVV-----SNGDVECTVVDKDCIITKIMKEADAVALDGLVYT 420
QY 428 AGKTKGLVPAAAGEHVPADSSN-----SYVAVVRRDSSHAFTLDELCKRKSCHAGF 480
DB 421 AG-VCGLVPVMAERYDDESQCKTDERPASFAVAVARKDSN--VWNWNLKGGKKSCHTAV 477
QY 481 GSPAGWDVPVVGALIQGFIRPKDCDLVAVSEFFNASCVPVNNPNKYPSLSCALCVGDRQ 540
DB 478 GRTAGWVPMG-LIHN---RTGTCN---FDEYFSEGCAV-GSPFN--SRLCQLCGSGG 526
QY 541 -GRNKCVCNSQERYGYGAFRCFLVENAGDVAFVHTTTFDNTNGHNSPEWAAELRSEY 599
DB 527 IPPEKCVASSHEKYEFGYTGALRCLEK-GDVAFIQHSTVEENTGGKKNADWAKNLQMDDF 585
QY 600 ELLCPNGARAEVSQFAACNLAIQIPPAVVRPDTNIFTVYGLLDKAQDLFG-DDRNKNGF 658
DB 586 ELLCTDGRANVMYRECNLAIEVTHAVVVRPE-KANKIRDLLERQEKFGVNGSEKSF 644
QY 659 KMFDSNYHGQDLFLFKDATVRAVPVGEKTYRGMGLDYVAALGEMSS-----QQCS 710
DB 645 WMFESQ---KDLFLKDLTKLKFVREGTYKFEFLGDKFYTVISSLKTCPNSDILQMCS 700
RESULT 13
A28438
N; Lactoferrin precursor - mouse
N; Alternate names: lactotransferrin
C; Species: Mus musculus (house mouse)
C; Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C; Accession: A28438; A41205
R; Pentecost, B.T.; Teng, C.T.
J. Biol. Chem. 262, 10134-10139, 1987
A; Title: Lactotransferrin is the major estrogen inducible protein of mouse uterine secre
A; Reference number: A92596; MUID:87280033; PMID:3611056
A; Accession: A28438
A; Molecule type: mRNA
A; Residues: 3-707 <PEN>
A; Cross-references: EMBL:J03298
R; Liu, Y.; Teng, C.T.
J. Biol. Chem. 266, 21880-21885, 1991
A; Title: Characterization of estrogen-responsive mouse lactoferrin promoter.
A; Reference number: A41205; MUID:92042099; PMID:1939212
A; Accession: A41205
A; Molecule type: DNA
A; Residues: 1-15 <LIU>
A; Cross-references: GB:M74778
C; Superfamily: transferrin; transferrin repeat homology
C; Keywords: duplication; glycoprotein
F; 1-19/Domain: signal sequence #status predicted <SIG>
F; 20-707/Product: lactotransferrin #status predicted <MAT>
F; 358-695/Domain: transferrin repeat homology <TRH2>
F; 494/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 30.3%; Score 1192.5; DB 1; Length 707;
Best Local Similarity 38.8%; Pred. No. 1.1e-82;
Matches 289; Conservative 116; Mismatches 254; Indels 85; Gaps 27;
QY 7 ALWILLALRLTVLGGWVVRWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRTSDADHCVQL 66
DB 13 ALGICLAKATT-----VFWCAVSNSEKCLRWQNMKRVG-GPPLSCVKKSSFTQCIOA 66
QY 67 TAAQEAADITLDGGAIEYAGK-EHGLKPVGVGV--DOEVGTSYAVAVVRRSSHVTIDT 123
DB 67 IVTNRADANTLDGGTMTFDAGKPPYKLRVAAEVYGTKEQPRTHYVAVVAVVKNSSNFHLNQ 126

A:Title: Transferrin messenger ribonucleic acid: molecular cloning and hormonal regulation
A:Reference number: A30014; MUID:87053639; PMID:3023031
A:Accession: A30014
A:Molecule type: mRNA
A:Residues: 518-687; 'D', '689-692', 'TA', '695 <HUG>
A:Cross-references: GB:M27966; NID:G207439; PIDN:AAA42267.1; PID:G207440
R:Schreiber, G.; Dryburgh, H.; Millership, A.; Maceuda, Y.; Inglis, A.; Phillips, J.; E. J. Biol. Chem. 254, 12013-12019, 1979
A:Title: The synthesis and secretion of rat transferrin.
A:Reference number: A14679; MUID:80049855; PMID:500689
A:Accession: A14679
A:Molecule type: protein
A:Residues: 20-47 <SCH>
R:Cavanaugh, P.G.; Nicolson, G.L. J. Cell. Biochem. 47, 261-271, 1991
A:Title: Lung-derived growth factor that stimulates the growth of lung-metastasizing tumor cells.
A:Reference number: A53289; MUID:92165927; PMID:1791188
A:Accession: A53289
A:Status: preliminary
A:Molecule type: protein
A:Residues: 89, 'Y', '91', 'A', '93-99', 'V', '101-102', 'N', '233', 'AN', '236-243', '401-406', 'N', '408 <CAV>
A:Experimental source: Lung
A:Note: sequence modified after extraction from NCBI backbone
A:Note: sequence extracted from NCBI backbone (NCBI:86115, NCBI:86116, NCBI:86114)
R:Purves, L.R.; Purves, M.; Linton, N.; Brandt, W.; Johnson, G.; Jacobs, P. Biochim. Biophys. Acta 966, 318-327, 1988
A:Title: Properties of the transferrin associated with rat intestinal mucosa.
A:Reference number: A30512; MUID:98327006; PMID:3046665
A:Accession: A30512
A:Molecule type: protein
A:Residues: 20-30; 619-643, 'KD', '646', 'LKACD', 'PUR>
R:Aldred, A.R.; Howlett, G.J.; Schreiber, G. Biochem. Biophys. Res. Commun. 122, 960-965, 1984
A:Title: Synthesis of rat transferrin in Escherichia coli containing a recombinant bacteriophage T4.
A:Reference number: I52203; MUID:84307580; PMID:6236811
A:Accession: I52203
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 7-25, 'X', '27-56', 'A', '58-64', '267-295 <RES>
A:Cross-references: GB:M26113; NID:G207437; PIDN:AAA42266.1; PID:G207438
C:Keywords: Duplication
A:Gene: TF
C:Superfamily: transferrin; transferrin repeat homology
F:20-348/Domain: transferrin repeat homology <TRH1>

Query Match 29.5%; Score 1160; DB 2; Length 695;
Best Local Similarity 38.8%; Pred. No. 3.1e-80;
Matches 289; Conservative 107; Mismatches 252; Indels 96; Gaps 31;

Qy 1 MRGPGALWLLALRTVLG--GMEVRWCATSDPEQKHC---GNMSEAPREAGIQSLLC 54
Db 1 MRFAVGALLACALGLCLAVPDKTVKCAVSEHNTKCI\$FRDHMTVL PADG--PRLPC 58
Qy 55 VRGTSADHCVQLIAQAEADAITLDGGAIEYAG-KEHGLKPVGVEYD--QEVGTSYYAVA 111
Db 59 VKKTSYQDCIKAI\$SGGEADAITLDGGWYDAGLTPNNLKFPVAAEFYGSLEHRTQHYLAVA 118
Qy 112 VWRSSHVITDITLKGKVSCHTGNRTVGNVNPVGYLVESGRSLVMGCD-----VLKAV 164
Db 119 VVKKGTDFQNLQKGGKSTGLGRAGMTIPTIGLFF-----CNLPEPRKPLEKAV 169
Qy 165 SDYFGGSCVPAGETSYSESLCLRCGDSGEGVCDSK\$PLERYDYDSGAPRCLAEAGDV 224
Db 170 ASFFSGSCVPADPVAFPQ-LCQLCPG-----CCSPTQPFPGYVGA\$FKLRDGGDV 221
Qy 225 AFVKGH\$TVLENTDGKTLPSWGQALLSQDFELLCRDGR\$ADVTWROCHLARVPAH\$VVVR 284
Db 222 AFVKHTTTFE----VLP---QKADRDQYELLCLDNTRKPEVDQYEDCYLARI\$PSHAVVAR 273
Qy 285 -ADTDGLI\$PRLLNEGQR\$F\$H-EGGSFQMF\$SEAYGQKDLL\$KDS\$TSELVPTATQTYEA 342
Db 274 NGDGKEDI\$WEILLKVAQ\$HFGK\$K\$DQ\$FUG\$PL--GKOLL\$KDSR\$FGLRAP\$KQGLQA 331

QY 343 -----WLGHEYLHAMKLLCDPNRLPPYLRWCVLSTPEIQKCGDMAVAFRRQRLKPEIQCV 398
Db 332 VPRPOLCH--CHSKSAGSCPDIDAIDSAPVKWICALSHOERAKCDEWSTNGQ-----IECE 384
QY 399 SAKSPQHCHMERIOAEQVDAVITLSGEDIYTAGKKYGLVPAAGBHYAPEDSSN-----S 450
Db 385 SAESTEDCIDKLVINGEADAMSLDGGHAYIAG--QGLVPMWENYDISSCTNPQSDVFPKG 443
QY 451 YVVAVVRDSSHAFTLDELROKRSCHAGFGSPAGWDVPVGALIQRGFIRPKDCDVLTA 510
Db 444 YVAVAVRASDS--SINMNNLKCKKSGCHTGVDRTAGWNPGLL-----FSRINHC-----KF 494
QY 511 SEFFNASCVPVNNPKNYPSSLCALCVGDEQGRNKCVCNSQERYYYGRGAFRCCLVENAGDV 570
Db 495 DEFFSQGCAP--GYKKR--STLCDLCIGPA-----KCAPNNREGYNGYTGAFCQCLVFK--GDV 546
QY 571 AFVRHTTVPDNTNGHNSPEWAEALRSEDYELLCPCNGARAEVSQFAACNLAQIPPHAVMVR 630
Db 547 AFVKQTVLENTNGKNTAAWKDLQEDFOLLCPDGTGKPKVTEFATCHLAQAPNHHVVS 606
QY 631 PD--TNIFTVYGLLPAQDLF--GDDHNKNGFKMFDSSNYHGQDLLFKDQATVRAVPVGEK 686
Db 607 KEKAARVSTV--LTAQKDLFWKDKDCTGNFCFLRSST---KDLLFRDDTKLTKLPEG 660
QY 687 TTYRGWGLDYVAALGHSQQCS 710
Db 661 TTYEYLGAEYLQAVGNI--RKCS 682

RESULT 15

S12100
transferrin precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Aug-1999
R:Moekaitis, J.E.; Pastori, R.L.; Schoenberg, D.R.
Nucleic Acids Res. 18, 6135, 1990
A:Title: The nucleotide sequence of xenopus laevis transferrin mRNA.
A:Reference number: S12100; MUID:91045087; PMID:2235503
A:Accession: S12100
A:Molecule type: mRNA
A:Residues: 1-717 <MOS>
A:Cross-references: EMBL:X54530; NID:G65158; PIDN:CAA38396.1; PID:G65159
C:Superfamily: transferrin; transferrin repeat homology
F:346-689/Domain: transferrin repeat homology <TRH2>

Query Match 28.7%; Score 1128.5; DB 2; Length 717;
Best Local Similarity 37.7%; Pred. No. 8.3e-78;
Matches 277; Conservative 106; Mismatches 252; Indels 99; Gaps 26;
QY 22 EYRWCATSDPEOHKCGNNSEAFREAGIQPSLLCVRGTSADHCVQLIAAQEAADAILDGA 81
Db 25 QVRWCVKNSSELKCKDLVDTCNKEIK--LSCVEKSNTECSLLFRKTMQMVFVMTGDD 82
QY 82 IYEAG--KEHGLKPVVGEVDQ--EYVTSYAVAVVRRSSHVTIDTLKGKVSCHTGINRTV 138
Db 83 VYKSLQPNLKPMAENYSGTSTETDTCYAVAVVKKSGSKFTFDELKOKKSCHTGIGKTA 142
QY 139 GNNVPVGYLVESGRSLVSMGCDV---LKAVSDYFGSCVPAGETSYSBSLCLRCRCRGS 195
Db 143 GNNIIGULLERKLLKWAGPDSFTWRNAVSKFKAFCVPGAKEPKLSQ-----LCAGIK-- 196
QY 196 EGVCDKSPLEYYDYSYGAFCRLAEGAGDVAFVKHSTVLENTDGTLPWNGQALLSQDFEL 255
Db 197 EHKSRSNNEPYNNYAGAFKCLQDDQGDVAFVKQSTVPEE-----FHKDYEL 243
QY 256 LCDGSRADVTBWRCHLARVPAHVAHVVRADTDGGL--IFRLINEGQ-----RLF-----SH 305
Db 244 LCPDNRKSIKEYKCNLAKVPAHVAHVLRGRDDKSKDIIIEFLQEAQKQCECKLRLPGWG 303
QY 306 EGSSQFMFSSEA-----YQKDLLFKDSTSELVPIATQTYEAWLGHEYLHAMK--GLLCD 358

Db 304 KGSNFQQRSESYSPPIFYQ-----FSVPSRSL-----FQCIQALKEGVKED 346
QY 359 PNRLPPYLRWCVLSTPEIQKCGDMAVAFRRQRLKPEIQCVSAKSPQHCHMERIOAEQVDAV 418
Db 347 DSAAQVKVRWCTQSKAERTKCDWTTI-----SGGAIECTEASTAEECIVQILKGDADAV 401
QY 419 TLSGEDIYTAGKKYGLVPAAGBHYAPED-----SSNSYVVAVVRDSSHAFTLD 468
Db 402 TLDGYSMTAG--LCGLVPMGEIYDQDDLTPCQSRSCSOAKGVYVAVAVIKKGTQVWS-- 458
QY 469 ELRGRKRSCHAGFGSPAGWDVPVGALIQRGFIRPKDCDVLTA VSEFFNASCVPVNNPKNY 528
Db 459 NLRGVKTCTAVGRTAGWNPGLITS-----ETANCDFAVYGE-----SCAPGSDVK-- 507
QY 529 SSLCALCVGD-----EQGRNKCVCNSQERYYYGRGAFRCCLVENAGDVAFVRHTTVPDNTNG 584
Db 508 SNLCALCIGDPEKLSEREKCKSPSASEAYYSGAFRCCLVEK--GOVGFAKHTTVPENTDG 566
QY 585 HNSEPMAAELRSEDYELLCPCNGARAEVSQFAACNLAQIPPHAVMVRPDTNIFTVYGLLDK 644
Db 567 KNPAGWAKDLKSEDFELLCPDGSRAPVTDYKRCNLAEVPAHVAHVTLTPDKR--EQVAKI VVN 625
QY 645 AQDLFG--DDHNKNGFKMFDSSNYHGQDLLFKDQATVRAVPVGEKTTYRGWGLDYVAALG 703
Db 626 QOSLYGRKGFQKIDIFQMFQSTG--GKDLLFKDSTOCLLEIPSKTTMQEFLGDKYHTAVTS 683
QY 704 MSSQOCSGAAAPAP 717
Db 684 LNKCSSTNEASWLP 697

Search completed: May 14, 2004, 09:43:58
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:36:02 ; Search time 17 Seconds
(without alignments)
2260.457 Million cell updates/sec

Title: US-10-049-957-4

Perfect score: 3936

Sequence: 1 MRGPSALWLLALRTVLGG.....APLLPLLALPALAARLLPPAL 738

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3936	100.0	738	1 TRFM HUMAN	P08582 homo sapien
2	3300	83.8	738	1 TRFM MOUSE	Q9r0r1 mus musculus
3	1308	33.2	690	1 TRFE ORYLA	P79819 oryzias lat
4	1274	32.4	691	1 TRF2 SALSA	P80429 salmo salar
5	1272.5	32.3	690	1 TRF1 SALSA	P80426 salmo salar
6	1269.5	32.3	698	1 TRFE HUMAN	P02787 homo sapien
7	1263.5	32.1	706	1 TRFE HORSE	P27425 equus caball
8	1261	32.0	695	1 TRFL HORSE	O77811 equus caball
9	1260.5	32.0	711	1 TRFL HUMAN	P02788 homo sapien
10	1258	32.0	708	1 TRFL CAPHI	Q29477 capra hircu
11	1249	31.7	708	1 TRFL BOVIN	P24627 bos taurus
12	1249	31.7	708	1 TRFL CAMDR	Q9tum0 camelus dro
13	1248	31.7	687	1 TRFE ONCKI	P79815 oncorhynchu
14	1240	31.5	695	1 TRFE RABIT	P19134 oryctolagus
15	1236.5	31.4	695	1 TRFE PAROL	O93429 paralichthy
16	1229	31.2	708	1 TRFL BUBBU	O77698 bubalus bub
17	1225.5	31.1	704	1 TRFE BOVIN	Q29443 bos taurus
18	1224.5	31.1	704	1 TRFL PIG	P14632 sus scrofa
19	1213.5	30.8	696	1 TRFE PIG	P09371 sus scrofa
20	1205.5	30.6	701	1 TRFE XENLA	P20333 xenopus lae
21	1202.5	30.6	704	1 ICA PIG	Q29545 sus scrofa
22	1202.5	30.6	705	1 TRFE CHICK	P02789 gallus gall
23	1199.5	30.5	707	1 TRFL MOUSE	P08071 mus musculus
24	1196	30.4	697	1 TRFE MOUSE	Q92111 mus musculus
25	1190.5	30.2	698	1 TRFE RAT	P12346 rattus norv
26	1163	29.5	642	1 TRFE GADMO	Q92079 gadus morhu
27	1156	29.4	686	1 TRFE ANAPL	P56410 anas platyr
28	964.5	24.5	844	1 SAX RANCA	P31226 rana catesb
29	824	20.9	726	1 TRF_BLADI	Q02942 blaberus di
30	551.5	14.0	681	1 TRF_MANSE	P22297 manduca sex
31	365	9.3	629	1 TRF_SARPE	Q26643 sarcophaga
32	126	3.2	1639	1 LMGI DROME	P15215 drosophila
33	111.5	2.8	640	1 RAEP_CANAL	O93831 candida alb

RESULT 1

ID	TRFM HUMAN	STANDARD;	PRT;	738 AA.
DT	01-AUG-1988 (Rel. 08, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Melanotransferrin precursor (Melanoma-associated antigen p97) (CD228 antigen)			
DE	antigen)			
GN	MF12 OR MAP97.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RC	TISSUE=Melanoma;			
RX	MEDLINE=86149285; PubMed=2419904;			
RA	Rose T.M., Plozman G.D., Teplow D.B., Dreyer W.J., Hellstroem K.E., Brown J.P.;			
RT	"Primary structure of the human melanoma-associated antigen p97 (melanotransferrin) deduced from the mRNA sequence.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 83:1261-1265(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM 2).			
RC	TISSUE=Skin, and Uterus;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Brownstein M.J., Udutin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smalish D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RN	[3]			
RP	GPI-ANCHOR.			
RX	MEDLINE=94132080; PubMed=8300636;			
RA	Food M.R., Rothenberger S., Gabathuler R., Haidl I.D., Reid G., Jefferies W.A.;			
RT	"Transport and expression in human melanomas of a transferrin-like glycosylphosphatidylinositol-anchored protein.";			
RL	J. Biol. Chem. 269:3034-3040(1994).			
RN	[4]			

34	111	2.8	474	1	GLGA_XANAC
35	110.5	2.8	419	1	PROA BORPE
36	107	2.7	515	1	PDI WHEAT
37	107	2.7	899	1	ZCH2 HUMAN
38	104.5	2.7	513	1	PDI HORVU
39	104	2.6	756	1	PDI MOUSE
40	102	2.6	1455	1	FACA HUMAN
41	99	2.5	929	1	SYA HALN1
42	99	2.5	3487	1	CSM2 HUMAN
43	98.5	2.5	655	1	HGFA HUMAN
44	97.5	2.5	650	1	GYBB_MYCPN
45	97.5	2.5	2511	1	FAS_CHECK

ALIGNMENTS

Q8pqa3	xanthomonas
Q7vwa0	bordetella
P52589	triticum ae
Q9cob9	homo sapien
P80384	hordeum vul
Q8r3b1	mus musculus
O15360	homo sapien
Q9hn24	halobacteri
Q04756	homo sapien
P22447	mycoplasma
P12276	gallus gall

RP FUNCTION.
 RX MEDLINE=96016189; PubMed=7556058;
 RA Kennard M.L., Richardson D.R., Gabathuler R., Ponka P.,
 RA Jefferies W.A.;
 RL "A novel iron uptake mechanism mediated by GPI-anchored human p97.";
 RT EMBO J. 14:4178-4186(1995).
 RN [5]
 RP IRON-BINDING.
 RX MEDLINE=92183868; PubMed=1544447;
 RA Baker E.N., Baker H.M., Smith C.A., Stebbins M.R., Kahn M.,
 RA Hellstroem K.E., Hellstroem I.;
 RT "Human melanotransferrin (p97) has only one functional iron-binding
 site.";
 RN FEBS Lett. 298:215-218(1992).
 RN [6]
 RP 3D-STRUCTURE MODELING.
 RX MEDLINE=92339524; PubMed=1633859;
 RA Garrat R.C., Jhoti H.;
 RT "A molecular model for the tumour-associated antigen, p97, suggests a
 Zn-binding function.";
 RL FEBS Lett. 305:55-61(1992).
 CC -!- FUNCTION: Involved in iron cellular uptake. Seems to be
 CC internalized and then recycled back to the cell membrane. Binds a
 CC single atom of iron per subunit. Could also bind zinc.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=P08582-1; Sequences=Displayed;
 CC Name=2;
 CC IsoId=P08582-2; Sequences=VSP 006557, VSP 006558;
 CC Notes-No experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Found predominantly in human melanomas and in
 CC certain fetal tissues; also found in liver, epithelium, umbilical
 CC chord, placenta and sweat gland ducts.
 CC -!- DOMAIN: Composed of two homologous domains.
 CC -!- SIMILARITY: Belongs to the transferrin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M12154; AA55992.1; -;
 DR EMBL; A00127; CAA00012.1; -;
 DR EMBL; BC001875; AAH01875.1; -;
 DR EMBL; BC002623; AAH02623.1; -;
 DR EMBL; BC007550; AAH07550.1; -;
 DR PIR; A23814; TFHUM.
 DR HSP; P19134; 1TFD.
 DR Genew; HGNC:7037; MF12.
 DR MIM; 155750; -;
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0005506; F:iron ion binding; TAS.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF04005; transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00307; TRANSFERRIN_3; 2.
 KW Transport; Iron transport; Glycoprotein; Lipoprotein; Metal-binding;
 KW Repeat; Signal; GPI-anchor; Membrane; Zinc; Alternative splicing.
 FT SIGNAL 1 19
 FT CHAIN 20 709 MELANOTRANSFERRIN (POTENTIAL).
 FT PROPEP 710 738 REMOVED IN NATURE FORM (POTENTIAL).
 FT REPEAT 20 361 1.
 FT REPEAT 362 713 2.
 FT DISULFID 26 63 BY SIMILARITY.
 FT DISULFID 36 54 BY SIMILARITY.

FT DISULFID	130	216	BY SIMILARITY.
FT DISULFID	172	189	BY SIMILARITY.
FT DISULFID	186	199	BY SIMILARITY.
FT DISULFID	257	271	BY SIMILARITY.
FT METAL	78	78	IRON 1 (BY SIMILARITY).
FT METAL	107	107	IRON 1 (BY SIMILARITY).
FT METAL	210	210	IRON 1 (BY SIMILARITY).
FT METAL	279	279	IRON 1 (BY SIMILARITY).
FT METAL	421	421	IRON 2 (BY SIMILARITY).
FT METAL	451	451	IRON 2 (BY SIMILARITY).
FT METAL	556	556	IRON 2 (BY SIMILARITY).
FT METAL	625	625	IRON 2 (BY SIMILARITY).
FT BINDING	132	132	CARBONATE 1 (BY SIMILARITY).
FT BINDING	136	136	CARBONATE 1 (BY SIMILARITY).
FT BINDING	138	138	CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING	139	139	CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT LIPID	709	709	GPI-anchor amidated cysteine (Potential). N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	38	38	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	135	135	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	515	515	N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC	238	302	KTLPSGQALLSQDFELLCRDGRADVTVMQCHLARVPA HAVVVRADTGGILFRLLNEGQL -> ESPSRQTWRSE EEEGCPAHEARTRMRSSAGQAKWAPVHRPQDESCKGEF GKRAKSRDMLG (in isoform 2). /FTId=VSP 006557. Missing (in isoform 2). /FTId=VSP 006558.
FT VARSPLIC	303	738	
FT SEQUENCE	738 AA; 80241 NW; 6E0868E894D7B955	CRC64;	
Query Match	100.0%; Score 3936; DB 1; Length 738;		
Best Local Similarity	100.0%; Pred. No. 1.7e-292;		
Matches 738; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY 1	MRGPGALWLLALRTVLGGMVWRWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTS	60	
DB 1	MRGPGALWLLALRTVLGGMVWRWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTS	60	
QY 61	DHCVOLIAAQADAITLGGAIYEAGKEHGLKPVVGEVYDQEVGTSYAVAVRRSSHT	120	
DB 61	DHCVOLIAAQADAITLGGAIYEAGKEHGLKPVVGEVYDQEVGTSYAVAVRRSSHT	120	
QY 121	IDTLKGVSCHTGINTRTGMVNPVGYLVESGRSLVMGCDVLKAVSDYFGGSCVPGAGTS	180	
DB 121	IDTLKGVSCHTGINTRTGMVNPVGYLVESGRSLVMGCDVLKAVSDYFGGSCVPGAGTS	180	
QY 181	YSESLCRLCRGDSGEGVCDKSPLEYYDYSGAFRCLEAGADVAFAVKHSTVLENTDGT	240	
DB 181	YSESLCRLCRGDSGEGVCDKSPLEYYDYSGAFRCLEAGADVAFAVKHSTVLENTDGT	240	
QY 241	LPSWQALLSQDFELLCRDGRADVTVMQCHLARVPAHAVVVRADTGGILFRLLNEQ	300	
DB 241	LPSWQALLSQDFELLCRDGRADVTVMQCHLARVPAHAVVVRADTGGILFRLLNEQ	300	
QY 301	RLFSHEGSSGFQMFSSSEAYGQDLDFKSTSELVPIATQTYEAWLGHEYLHAKGLLCDN	360	
DB 301	RLFSHEGSSGFQMFSSSEAYGQDLDFKSTSELVPIATQTYEAWLGHEYLHAKGLLCDN	360	
QY 361	RLPYLRWCVLSTPEIQKCGDMAVAFRRQRKPEIQCVSAKSPOHCMERIOAEQDVAVL	420	
DB 361	RLPYLRWCVLSTPEIQKCGDMAVAFRRQRKPEIQCVSAKSPOHCMERIOAEQDVAVL	420	
QY 421	SGEDIYTAGKYGLVPAAGEHYAPEDSSNSYVVAVVRDSSHAFTLDELGRKRSCHAGF	480	
DB 421	SGEDIYTAGKYGLVPAAGEHYAPEDSSNSYVVAVVRDSSHAFTLDELGRKRSCHAGF	480	
QY 481	GSPAGWDVPVCGALIORGIRPKDCDLTAVSEFFNASCVPVNNPKNYPSSLCALCVGBQ	540	
DB 481	GSPAGWDVPVCGALIORGIRPKDCDLTAVSEFFNASCVPVNNPKNYPSSLCALCVGBQ	540	
QY 541	GRNCKVGNQSBRYGCGAFRCLENACDVAFAVRHTTTFDNTNGNSPFWAAELSEDEYE	600	
DB 541	GRNCKVGNQSBRYGCGAFRCLENACDVAFAVRHTTTFDNTNGNSPFWAAELSEDEYE	600	

Db 1 MRLSVTFLLSLRTVVCVMEVQWCTISDAEQKCKMDSEAFQAGIRPSLLCQGN5A 60
 Qy 61 DHCYQLIAQAEADAITLDGAIYEAGKEHGLKPVVGEVYDQVCTSYAYAVVRRSSHVT 120
 Db 61 DHCYQLIKEQADAITLDGAIYEAGKEHGLKPVVGEVYDQVCTSYAYAVVRRSNVT 120
 Qy 121 IDTLKGKCHTGINRTKWNVPVGLVBSGRSLVWGCDVLKAVSDYFGSGCPVAGETS 180
 Db 121 INTLKGKCHTGINRTKWNVPVGLVBSGRSLVWGCDVLKAVSDYFGSGCPVAGETS 180
 Qy 181 YSESLCRLCRGSSGEGVCDKSPLEYYDYSAGFRLAEGAGDVAFVKGHSTVLENTDGT 240
 Db 181 HSESLCRLCRGSSGEGVCDKSPLEYYDYSAGFRLAEGAGDVAFVKGHSTVLENTDGT 240
 Qy 241 LPSWGQALLSQPELLCRGSRADVTWQCHLARVPAAHVAVVRADTDGGLIFRLNEQ 300
 Db 241 LPSWGKSLMSDFOLLCRGSRADITWRRCHLAKVPAHVAVVRGMDGGLIFOLLNEQ 300
 Qy 301 RLFSHEGSSFFQSSKAYSKQKLLFKDSTSELVPIATQTYEAWLGHVHLHMKGLLCPN 360
 Db 301 RLFSHEDSSFFQSSKAYSKQKLLFKDSTSELVPIATQTYEAWLGHVHLHMKGLLCPN 360
 Qy 361 RLPLYLRCWLSTPEIKQKGMMAVAFRRQRLKPEIQCVSAKSPQHCMERIQAEQVDAVTL 420
 Db 361 RLPLYLRCWLSTPEIKQKGMMAVAFRRQRLKPEIQCVSAKSPQHCMERIQAEQVDAVTL 420
 Qy 421 SEDITYAGKTKGLVPAAGEHAPEDSSNSYVAVVRDSSHAFTLDELGRKRSCHAGF 480
 Db 421 RGEDYRAGKTKGLVPAAGEHAPEDSSNSYVAVVRDSSHAFTLDELGRKRSCHAGF 480
 Qy 481 GSPAGWDVPVGLIQRGFRPKDCDVLTAVSFFFNASCVPVNNKPNYPSLICALCVGDEO 540
 Db 481 GSPAGWEVPIGSLIQRGFRPKDCDVLTAVSFFFNASCVPVNNKPNYPSLICALCVGDEK 540
 Qy 541 GRNKCVCNSQERYGYRGAFRLCVENAGDVAFVRRHTTVDNTNGHNSPAAELRSEDEY 600
 Db 541 GRNKCVCNSQERYGYRGAFRLCVENAGDVAFVRRHTTVDNTNGHNSPAAELRSEDEY 600
 Qy 601 LLCPNGAREVSOFAACNLAIQIPPHAVMVRPDTNIFTVYGLLDKADLFGDDHKNKGFQ 660
 Db 601 LLCPNGAREVSOFAACNLAIQIPPHAVMVRPDTNIFTVYGLLDKADLFGDDHKNKGFQ 660
 Qy 661 FDSSNYHGDLDFKATRAVPVGEKTYRGWGLDYVAALGCMSSQCSGAAAPAGAP 720
 Db 661 FDSSNYHGDLDFKATRAVPVGEKTYRGWGLDYVAALGCMSSQCSGAAAPAGAP 720
 Qy 721 LLPLLLPALAARLLPPAL 738
 Db 721 LLALLLLTLAGLLPRVL 738

RESULT 3

ID TRFE OXYLA STANDARD; PRT; 690 AA.
 AC P79619;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Serotransferrin precursor.
 OS Oryzias latipes (Medaka fish) (Japanese ricefish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
 OX NCBI_TaxID=8090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96414927; PubMed=8817928;
 RA Mikawa N., Hirano I., Aoki T.;
 RT "Structure of medaka transferrin gene and its 5'-flanking region."
 RL Mol. Mar. Biol. Biotechnol. 5:225-229 (1996).
 CC -!- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING

CC OF AN ANION, USUALLY BICARBONATE.
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- DOMAIN: Composed of two homologous domains.
 CC -!- SIMILARITY: Belongs to the transferrin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D64033; BAAL0901.1; -.
 CC HSP; P56410; IAOV.
 CC InterPro; IPR001156; Transferrin.
 CC Pfam; PF00405; transferrin; 2.
 CC PRINTS; PR00422; TRANSFERRIN.
 CC SMART; SM00094; TR_PER; 2.
 CC PROSITE; PS00205; TRANSFERRIN_1; 2.
 CC PROSITE; PS00206; TRANSFERRIN_2; 2.
 CC PROSITE; PS00207; TRANSFERRIN_3; FALSE NEG.
 CC Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 CC Signal.
 CC
 CC CHAIN 1 17 BY SIMILARITY.
 CC SEROTRANSFERRIN.
 CC DISULFID 18 690 BY SIMILARITY.
 CC DISULFID 27 49 BY SIMILARITY.
 CC DISULFID 126 206 BY SIMILARITY.
 CC DISULFID 171 185 BY SIMILARITY.
 CC DISULFID 234 248 BY SIMILARITY.
 CC DISULFID 343 379 BY SIMILARITY.
 CC DISULFID 353 370 BY SIMILARITY.
 CC DISULFID 404 682 BY SIMILARITY.
 CC DISULFID 419 643 BY SIMILARITY.
 CC DISULFID 452 530 BY SIMILARITY.
 CC DISULFID 476 671 BY SIMILARITY.
 CC DISULFID 486 499 BY SIMILARITY.
 CC DISULFID 496 513 BY SIMILARITY.
 CC DISULFID 570 584 BY SIMILARITY.
 CC METAL 73 73 IRON 1 (BY SIMILARITY).
 CC METAL 103 103 IRON 1 (BY SIMILARITY).
 CC METAL 200 200 IRON 1 (BY SIMILARITY).
 CC METAL 256 256 IRON 1 (BY SIMILARITY).
 CC METAL 394 394 IRON 2 (BY SIMILARITY).
 CC METAL 429 429 IRON 2 (BY SIMILARITY).
 CC METAL 524 524 IRON 2 (BY SIMILARITY).
 CC METAL 592 592 IRON 2 (BY SIMILARITY).
 CC BINDING 128 128 CARBONATE 1 (BY SIMILARITY).
 CC BINDING 132 132 CARBONATE 1 (BY SIMILARITY).
 CC BINDING 134 134 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
 CC SIMILARITY).
 CC BINDING 135 135 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
 CC SIMILARITY).
 CC BINDING 454 454 CARBONATE 2 (BY SIMILARITY).
 CC BINDING 458 458 CARBONATE 2 (BY SIMILARITY).
 CC BINDING 460 460 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 CC SIMILARITY).
 CC BINDING 461 461 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 CC SIMILARITY).
 CC SEQUENCE 690 AA; 74600 MW; 2F996CALAE79570 CRC64;
 Query Match 33.2%; Score 1308; DB 1; Length 690;
 Best Local Similarity 40.3%; Pred. No. 5.8e-92;
 Matches 300; Conservative 116; Mismatches 215; Indels 114; Gaps 29;
 Qy 12 LALRTVLG-----GMEVRCWCATSDPEQHKCNMSEAFREAGIQPSLLCVRGTSADH 62
 Db 4 LLLLTLLCLLAALAVPAQKWKVCQDFEKKCSDLAA-----SFAFSCVKKESTLD 57
 Qy 63 CVQLIAQAEADAITLDGAIYEAG-KEHGLKPVVGEVYDQVCTSYAYAVVRRSSHVT 121
 Db 58 CIIAIKAGEADAITVDGGDVYTAGLNNDLHPHIAEDYGTSETCYAVAVAKGTTFGI 117


```
QY 120 TIDTLKGVKSCHTGINRTVGMNPGVYLVSGLSMVGC---VLKAVSDVFGGSCVPGA 176
DB 117 GFKTLRGKSKHTGLGKAGWNIPITGLVTESQIRKAGIEDRPVESAVDSDFNASCAPGA 176
QY 177 GETSYSESLRCLRGDSGEGVCDKSPLEYYDYSGAFRCIAEAGDGVAFVKHSTVLENT 236
DB 177 ---TWGSKLCLQCKGD-----CSRSHKEPYDYAGAFQCLKDGAGDGVAFIKPLAV--- 223
QY 237 DGKTLPSWGQALLSDPFLLCRDGRADVTEWRQCHLARVPAHVVVVVADTDGGLIFRL 296
DB 224 -----PRAEKASYELLCKDGTFRASIDSYKTCHLARVPAHVVVSRDPE--LANRIY 272
QY 297 NEGQRLFSEHGSFOWFSEFAGVQKDLFLKDSSTSELVPIATOTYE-AWLGHEYLHAMKGL 355
DB 273 NKLAV-----KDFLNFSDGVAANLWFKQSAQKLQVLPPTTDSFLYLGAEYMTIRSL 327
QY 356 LCD--PNRLPPYLRVCLSTPEIQCGDMV-AFRRLKPEIQCVSAKSPQHCHMERIOA 412
DB 328 KKSQATGASSRAIKWCAVGAHEKGCDDTWTINSFADGESK--ISCQDAPTVEECIKIMR 385
QY 413 EQVDVAVTSGEDIYTAGKYGLVPAAGEHY-----APEDSNSYVVAVVRDSSHAFT 466
DB 386 KEADIAVVDGGEVYTAG-KCGLVPVWVEQYDADLCSAPCEAS-SYAVAVAKKGS--GLT 441
QY 467 LDELGRKSGHAGFGSPAGDVPVPCALIQRGFIRPKDCDVLTVAVSEFFENASCVP---VNN 523
DB 442 WTKLKGKRSCHTGLGRTAGWNIPMGLIHQ-----ETNDCD----FTKYFSKGCAPGSEVGS 493
QY 524 PKNYPSSLCALCVGDEQGR---NKCVGNSQBYRYGYRGAFCLVLENAGDVAFVRHTTVF 579
DB 494 P-----FCAQCKGSGKARGGDEDRCKARSEQYIYGYTGAFCLVEDAGDVAFIKHTIVP 547
QY 580 DNTNGHSEPMAEELRSEDYELLCNGARAEVQFAACNLAQIPPHAVVVRPDNTIFTY 639
DB 548 ESTDG-NGPDWAKDLKSSDFELLQDGTTPVTKESECHLAKVPAHAVITRPTETR-GDVV 605
QY 640 GLLDRAQDLFGDHNKNGFKMFDSSNYHGODLLFKDAT--VRAVPVGEKTVYRGWGLGDY 697
DB 606 SILLELQAKFGSGSDSFRMFQSS--VERNLKFDSTKCLQEIPIKG--TKYQDFLGKEY 661
QY 698 VAALEGMSQQCSGAAA 714
DB 662 MIAMQ--SLRKCSDSSTS 676

RESULT 5
ID TRF1_SALSA STANDARD; PRT; 690 AA.
AC P80426.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Serotransferrin I precursor (Siderophilin I) (STF I).
GN STF1.
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]_TaxID=8030;
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94122797; PubMed=8293074;
RA Kvingedal A.M., Roervik K.A., Alestrom P.;
RT "Cloning and characterization of Atlantic salmon (Salmo salar) serum
transferrin cDNA";
RL Mol. Mar. Biol. Biotechnol. 2:233-238 (1993).
RN [2]
RP SEQUENCE OF 1-64 FROM N.A.
RX MEDLINE=95121925; PubMed=7821802;
RA Kvingedal A.M.;
RT "Characterization of the 5' region of the Atlantic salmon (Salmo
```

```
RT salar) transferrin-encoding gene.";
RL Gene 150:335-339 (1994).
RN [3]
RP SEQUENCE OF 19-37.
RC TISSUE=Serum;
RA Roed K.H., Dehli A.K., Flengsrud R., Midtjell L., Roervik K.A.;
RT "Immunosay and partial characterization of serum transferrin from
Fish Shellfish Immunol. 5:71-80 (1995).
RL Fish Shellfish Immunol. 5:71-80 (1995).
CC -!- FUNCTION: Transferrins are iron binding transport proteins which
can bind two atoms of ferric iron in association with the binding
of an anion, usually bicarbonate. It is responsible for the
transport of iron from sites of absorption and heme degradation to
those of storage and utilization. Serum transferrin may also have
a further role in stimulating cell proliferation.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Abundant in liver and serum with smaller
amounts found in the stomach and kidney.
CC -!- DOMAIN: Composed of two homologous domains.
CC -!- SIMILARITY: Belongs to the transferrin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
CC
CC -----
EMBL; L20313; AAA18838.1; -.
DR EMBL; L26909; AAC42221.1; -.
DR PIR; I51350; I51350.
DR PIR; T11749; T11749.
DR HSSP; P56410; IAOV.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00094; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 1.
KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
KW Signal.
FT SIGNAL 1 18
FT CHAIN 19 690 SEROTRANSFERRIN I.
FT DISULFID 28 50 BY SIMILARITY.
FT DISULFID 127 207 BY SIMILARITY.
FT DISULFID 172 186 BY SIMILARITY.
FT DISULFID 235 249 BY SIMILARITY.
FT DISULFID 343 379 BY SIMILARITY.
FT DISULFID 353 370 BY SIMILARITY.
FT DISULFID 404 681 BY SIMILARITY.
FT DISULFID 419 642 BY SIMILARITY.
FT DISULFID 451 529 BY SIMILARITY.
FT DISULFID 475 670 BY SIMILARITY.
FT DISULFID 485 499 BY SIMILARITY.
FT DISULFID 496 512 BY SIMILARITY.
FT DISULFID 569 583 BY SIMILARITY.
FT METAL 74 74 IRON 1 (BY SIMILARITY).
FT METAL 104 104 IRON 1 (BY SIMILARITY).
FT METAL 201 201 IRON 1 (BY SIMILARITY).
FT METAL 257 257 IRON 1 (BY SIMILARITY).
FT METAL 394 394 IRON 2 (BY SIMILARITY).
FT METAL 428 428 IRON 2 (BY SIMILARITY).
FT METAL 523 523 IRON 2 (BY SIMILARITY).
FT METAL 591 591 IRON 2 (BY SIMILARITY).
FT CARBOHYD 169 169 N-LINKED (GLCNAC...) (POTENTIAL).
FT BINDING 129 129 CARBONATE 1 (BY SIMILARITY).
FT BINDING 134 134 CARBONATE 1 (BY SIMILARITY).
FT BINDING 136 136 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
SIMILARITY).
FT BINDING 137 137 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
```

```
FT BINDING 453 453 SIMILARITY).
FT BINDING 457 457 CARBONATE 2 (BY SIMILARITY).
FT BINDING 459 459 CARBONATE 2 (BY SIMILARITY).
FT BINDING 460 460 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
FT BINDING 460 460 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
FT CONFLICT 34 34 SIMILARITY).
FT SEQUENCE 690 AA; 74597 MW; D1F4FC74A6AFA3FB CRC64;

Query Match 32.3%; Score 1272.5; DB 1; Length 690;
Best Local Similarity 39.1%; Pred. No. 3e-89;
Matches 289; Conservative 128; Mismatches 221; Indels 101; Gaps 29;

QY 11 LLALRTVLGME-----VRWCATSDPEQHKCGNMSEAFREAGIOPSLLCVRGTS 60
DB 3 LLLLSALLGLATAYAAPAGIVKVKSEQLKCHDLAAKVAE-----FSCVRKDG 56
QY 61 DHCVQIIAQEADAITLDGAIYAG--KEHGLKPVVGEVVDQVGTSGSYAVAVRRSSH 119
DB 57 FECIOAKGGEADAITLDGDIYTAGLNYGLOPIIAEDYGEDSDTCYAVAVAKKGTAF 116
QY 120 TIDTLKGVKSGHNGINRTWGVNVPVGLVSGRLSVMGCD--VLKAVSDYFGSCVPGA 176
DB 117 GFKTLRGKKSCHTGLGKSGAGNIPIGTLVTESQIRWAGIEDRPVESAVSDFNASC 176
QY 177 GETSYSESICRLCRGDSGBGVCDKSPLERYDYGAFRCIARAGDGVAFVXHSHTYLE 236
DB 177 ---TWGSKULCQCKGD-----CSRSHKEFYDYAGAFQCKDGAGDVAFIKPLAV 223
QY 237 DGKTLPSWGOALLSODPELLCRDGRADYVTEMQCHLARVPAHVVVRADTDGGLIFRL 296
DB 224 -----PAEBAKASYELLCKDGTASIDSYKNTCHLARVPAHVVVRADTDGGLIF 272
QY 297 NEGQRLFSHEGSGFQFSEAYCQKDLLFKDSTSELVPIATQTYE--AWLGEVYHAKGL 355
DB 273 NKLAV-----KDFNLFSSDGYAAKLMFKDSQAQLVQLPTTTDSFLYLGAETMSTIR 327
QY 356 LCD--PNRLPPYLRWCVLPSTPEIQKCDMAV-AFRQRRLKPEIQCVSAKSPQHCMEIO 412
DB 328 KKSQATGASSRAIKCAVGAHAEKGCDDTWTINSFADGESK--ISCQDAPTEVEICIK 385
QY 413 EQVDVTLGSDIYTAGKYGLVPAAGEHY-----APEDSSNSYVYVAVVRDSSHAFT 466
DB 386 KEADAIAVDGVEYTAG-KCGLVPVMVEQYDADLCSAPGEAS--SYTAVAVAKKGS-- 441
QY 467 LDELKGRSCHAGFGSPAGHDVPVGLIQRGFRPKDCDVLTVAFSEFFNASCVP----- 523
DB 442 WKTLLKGRSCHTGLGRTAGNINPMGLIHQ-----ETNDCD-----FTKYFGKGC 493
QY 524 PKNYPSSLCAIC-----VGDEGRNKCVCNSOERYGYRGAPRCLVENAGDVAFVRHT 577
DB 494 P-----FCAQCKGSKAVGDEY---RCRKARSEQYGYTGAPRCLVEDAGDGVAFI 544
QY 578 VFDNTNGHNSPWAELRSEDYELLCPNGARAEVSOFAACNLAQIPPHAVMVRPDTNIFT 637
DB 545 VPBSTDG-NGPDWAKDKSSDFELLCDQGTQTPVTKFSECHLAKVPAHAVITRPETR- 602
QY 638 VYGLLDKADQLFDDHKNKGKFMFSDSNVHGQDLLEFKDAT--VRAPVGEKTYRQWLGL 695
DB 603 VVSIILELOAKFGSSGSDSFRMFQSS--VEKNLLPKDSTKCIQETPKG--TKYQDFLGK 658
QY 696 DYVAALLEGHSSQOCGAAA 714
DB 659 EYMIAMQ--SLRKCSDSTS 675
```

RESULT 6

```
TRFE HUMAN STANDARD; PRT; 698 AA.
ID TRFE HUMAN
AC P02787; O43890; Q9NQ88; Q9UHV0;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
```

```
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal
GN binding globulin) (PRO1400).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A., AND VARIANTS TF*B2; TF*CHI AND TF*DI.
RX MEDLINE=84194084; PubMed=6585826;
RA van Bragt P.H., Baldwin W.D., Bowman B.H.;
RA "Human transferrin: cDNA characterization and chromosomal
RT localization."
RT Proc. Natl. Acad. Sci. U.S.A. 81:2752-2756(1984).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=88056305; PubMed=3678832;
RA Schaeffer E., Lucero M.A., Jeltsch J.-M., Py M.-C., Levin M.J.,
RA Chambon P., Cohen G.N., Zakim M.M.;
RT "Complete structure of the human transferrin gene. Comparison with
RT analogous chicken gene and human pseudogene."
RL Gene 56:109-116(1987).
RN [3]
SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92321139; PubMed=1809186;
RA Hershsberger C.B., Larson J.L., Arnold B., Rosteck P.R. Jr.,
RA Williams P., Dehoff B., Dunn P., O'Neal K.L., Riemen M.W.,
RA Tice P.A.;
RT "A cloned gene for human transferrin."
RL Ann. N.Y. Acad. Sci. 646:140-154(1991).
RN [4]
SEQUENCE FROM N.A., AND VARIANT ATRANSFERRINEMIA PRO-477.
RX MEDLINE=20563920; PubMed=11110675;
RA Beutler E., Gelbart T., Lee P.L., Trevino R., Fernandez M.A.,
RA Fairbanks V.F.;
RT "Molecular characterization of a case of attransferrinemia."
RL Blood 96:4071-4074(2000).
RN [5]
SEQUENCE OF 99-698 FROM N.A.
RC TISSUE=Fetal liver;
RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Bi J., Zhang Y., Liu M.,
RA He F.;
RT "Functional prediction of the coding sequences of 33 new genes deduced
RT by analysis of cDNA clones from human fetal liver."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
SEQUENCE OF 422-698 FROM N.A.
RX MEDLINE=84153910; PubMed=6322780;
RA Uzan G., Frain M., Park I., Besmond C., Maessen G., Trepas J.S.,
RA Zakim M.M., Kahn A.;
RT "Molecular cloning and sequence analysis of cDNA for human
RT transferrin."
RN [7]
SEQUENCE OF 20-698.
RX MEDLINE=83160878; PubMed=6833213;
RA McGilivray R.T.A., Mendez E., Shewale J.G., Sinha S.K.,
RA Lineback-Zins J., Brew K.;
RT "The primary structure of human serum transferrin. The structures of
RT seven cyanogen bromide fragments and the assembly of the complete
RT structure."
RN J. Biol. Chem. 258:3543-3553(1983).
RN [8]
SEQUENCE OF 73-698 FROM N.A.
RX MEDLINE=85216459; PubMed=3858812;
RA Park I., Schaeffer E., Sidoli A., Baralle F.E., Cohen G.N.,
RA Zakim M.M.;
RT "Organization of the human transferrin gene: direct evidence that it
RT originated by gene duplication."
RL Proc. Natl. Acad. Sci. U.S.A. 82:3149-3153(1985).
```


Db 220 DVAFVHKSTIFENLANKA-----DRDQYELLCLDNTRKPVDBYKDCHELAQVPSHTV 270
 QY 282 VVRADTDGG---LIFRLNQGRLFSHEGS-SFQMFSSSEAYGOKOLLFKDSTSELVPIAT 337
 Db 271 VAR---SMGKEDLIWELNQAEHFQKDKSEKFLFSS-PHG-KDLLFKDSAHGFLKVP 326
 QY 338 Q-TYEAWLGHYELHAKGLL-----CDPNRLPPYLRCVCLSTPEIKQCGDMAVA.385
 Db 327 RMDAKMYLGYEYVTAIRNLREGTCPEAPTDCKP-----VKWCALSHHERLKCDEWSV- 379
 QY 386 FRQRLKPEIQCSAKSPQCHMERIOAEQVDATVLSGEDITYAGKYGLVPAAGEHYAPE 445
 Db 380 ----NSVGKIECVASSTEDCTAKIMNGEADMSLDGGFYTAG-KCGLVPVLAENYKNS 434
 QY 446 DGSN-----SYVVAVVRDSSHAFTLDELGRKRSCHAGFSGPAGWDVPVGVGALIQGFIR 500
 Db 435 DCEPTPEAGYPAVAVVKSASD-LTWDNLGKKKSCHTAVGRTAGWNIPMGLL-----YNK 489
 QY 501 PKDCDVLTVASEFPNASCVPVNNPKYPSLSLCAVGDGQGRNKCVCNQSERYGYRGAP 560
 Db 490 INHC----RFDEFFSEGCAPGSKK---DSSLCKLCMG--SGLNLCEPNNKEGYGYTGAF 540
 QY 561 RCLVENAGDAVAFVRHTTVDNTNGHNSPEWAAELRSEDYELLCPCNGARAEVSQFAACNLA 620
 Db 541 RCLVEK-GDAFVKHQTVPQNTGGKPDPAWKNLNEKOEYELLCDDGTRKPVBEYANCHLA 599
 QY 621 QIPPHAVMVRPDTNITFTVYGLLDKAQDLFGDDHN--KNGFKMPSDNYHGDQLLFKDATV 678
 Db 600 RAPNAVVRKDKKEA-CVHKILRQOHLFGSNVTDGSGNFCFLRSET---KOLLFRDDIV 655
 QY 679 RAVPVGKTYRQWGLDYYVAALGMSOOCs 710
 Db 656 CLAKLHNRNTEKYLGEEYVKAAGNL--RKCS 685

RESULT 7

TRFE_HORSE STANDARD; PRT; 706 AA.
 AC P27425;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Serotransferrin precursor (Transferrin) (Siderophilin) (Beta-1-metal binding globulin).
 GN TF.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 OX NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93277958; PubMed=8504171;
 RA Carpenter M.A., Broad T.E.;
 RT "The cDNA sequence of horse transferrin."
 RL Biochim. Biophys. Acta 1173:230-232(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Extraembryonic tissue;
 RA McDowell K.J., Adams M.H., Baker C.B.;
 CC Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Transferrins are iron binding transport proteins which can bind two atoms of ferric iron in association with the binding of an anion, usually bicarbonate. It is responsible for the transport of iron from sites of absorption and heme degradation to those of storage and utilization. Serum transferrin may also have a further role in stimulating cell proliferation.
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
 CC -!- DOMAIN: Composed of two homologous domains.
 CC -!- SIMILARITY: Belongs to the transferrin family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M69020; AAA30958.1; -;
 DR EMBL; U21127; AAA63684.1; -;
 DR PIR; S33761; S33761.
 DR HSSP; P02787; LA8E.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; P00422; TRANSFERRIN.
 DR SMART; SM00094; TR FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 2.
 KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 KW Signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 706 SEROTRANSFERRIN.
 FT REPEAT 20 357 1.
 FT REPEAT 358 706 2.
 FT DISULFID 26 64 BY SIMILARITY.
 FT DISULFID 36 55 BY SIMILARITY.
 FT DISULFID 134 215 BY SIMILARITY.
 FT DISULFID 174 190 BY SIMILARITY.
 FT DISULFID 177 198 BY SIMILARITY.
 FT DISULFID 187 200 BY SIMILARITY.
 FT DISULFID 248 262 BY SIMILARITY.
 FT DISULFID 360 623 BY SIMILARITY.
 FT DISULFID 366 398 BY SIMILARITY.
 FT DISULFID 376 389 BY SIMILARITY.
 FT DISULFID 423 701 BY SIMILARITY.
 FT DISULFID 441 684 BY SIMILARITY.
 FT DISULFID 474 550 BY SIMILARITY.
 FT DISULFID 498 692 BY SIMILARITY.
 FT DISULFID 508 522 BY SIMILARITY.
 FT DISULFID 519 533 BY SIMILARITY.
 FT DISULFID 590 604 BY SIMILARITY.
 FT DISULFID 642 647 BY SIMILARITY.
 FT METAL 79 79 BY SIMILARITY.
 FT METAL 111 111 IRON 1 (BY SIMILARITY).
 FT METAL 209 209 IRON 1 (BY SIMILARITY).
 FT METAL 270 270 IRON 1 (BY SIMILARITY).
 FT METAL 413 413 IRON 2 (BY SIMILARITY).
 FT METAL 449 449 IRON 2 (BY SIMILARITY).
 FT METAL 544 544 IRON 2 (BY SIMILARITY).
 FT METAL 612 612 IRON 2 (BY SIMILARITY).
 FT BINDING 136 136 CARBONATE 1 (BY SIMILARITY).
 FT BINDING 140 140 CARBONATE 1 (BY SIMILARITY).
 FT BINDING 142 142 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
 FT BINDING 143 143 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
 FT BINDING 476 476 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 480 480 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 482 482 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
 FT BINDING 483 483 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
 FT CARBOHYD 515 515 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 706 AA; 78094 MW; 1A0FA566C0409D8A CRC64;
 Query Match 32.1%; Score 1263.5; DB 1; Length 706;
 Best Local Similarity 41.2%; Pred. No. 1.5e-88;
 Matches 307; Conservative 113; Mismatches 237; Indels 89; Gaps 30;
 QY 1 MEGPSGALWLLALRTVLGGMEVRWCATSDPEQHKCNSEAFREAGIQPSLL-CVYRGTS 59
 Db 1 MRLAIRALLACAVLGLCLAEQTVRMCTVSNHVSFKASFRDSMKSVTPAPPLVACVKRTS 60

QY 540 QGRNKCVCNGBQRYGYRGAPCLVFNAGDVAFVRHTTVDNTNGHNSPWAELRSBDY 599
 Db ENENKCMPSNEERYGYGTGAFCLAEKAGDVAFVQDVTVLQNTDCKNSPWAELKQEDF 575
 QY 600 ELLCPNGARAEVSQFAACNLAIQIPHAVVRPDTHFTFYGLLDKXQDLFGDDHNG-- 657
 Db 576 ELLCLDGTGRKPVAAEASCHLAPNAPVAVVQSQD-RAQHLKKVFLQDQDFGG----NGPD 630
 QY 658 ----FKMFDSSNVHQDILLFKDQAVRVPVGVGKTTYRGWLGIDYVAALGMSQOCSCGAA 713
 Db 631 CPKCFCLFSET---KNLLFNQNTCLAELOQKTTTYEOLGSEYVTSITNL--RRCS--- 682
 QY 714 APAPGAPLL 722
 Db 683 ----SSPLL 687

RESULT 9
 TRFL HUMAN
 ID TRFL HUMAN STANDARD; PRT; 711 AA.
 AC P02788; O00756; Q16780; Q16785; Q16786; Q16789; Q96KZ4; Q96KZ5;
 AC Q9H1Z3;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Lactotransferrin precursor (Lactoferrin) [Contains: Lactoferrin A;
 DE Lactoferrin B; Lactoferrin C].
 GN LTF OR LF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RX MEDLINE=90384839; PubMed=2402455;
 RA Rey M.W., Woloshuk S.L., de Boer H.A., Pieper F.R.;
 RT "Complete nucleotide sequence of human mammary gland lactoferrin.";
 RL Nucleic Acids Res. 18:5288-5288(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Cho Y.Y.;
 RT Thesis (1994), Genetic Engineering Research Institute / Taejon, Korea.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Connely O.M.;
 RT Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Liang Q., Jimenez-Flores R., Richardson T.;
 RT "Molecular cloning and sequence analysis of human lactoferrin.";
 RL Submitted (DEC-1991) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RA Wei X., Han J., Rado T.A.;
 RT "Human neutrophil lactoferrin coding and 5' flanking region DNA
 sequences.";
 RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Cheng H., Chen X., Huan L.;
 RT "cDNA cloning and sequence analysis of human lactoferrin.";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [8]
 RP SEQUENCE OF 3-711 FROM N.A.
 RC TISSUE=Mammary gland;
 RX MEDLINE=90326549; PubMed=2374734;
 RA Powell M.J., Ogden J.E.;
 RT "Nucleotide sequence of human lactoferrin cDNA.";
 RL Nucleic Acids Res. 18:4013-4013(1990).
 RN [9]
 RP SEQUENCE OF 20-711.
 RX MEDLINE=85076667; PubMed=6510420;
 RA Metz-Boutigue M.-H., Jolles J., Mazurier J., Schoentgen F.,
 RA Legrand D., Spik G., Montreuil J., Jolles P.;
 RT "Human lactotransferrin: amino acid sequence and structural
 comparisons with other transferrins.";
 RL Eur. J. Biochem. 145:659-666(1984).
 RN [10]
 RP PRELIMINARY SEQUENCE OF 20-73; 134-171; 257-278 AND 347-530.
 RX MEDLINE=82046817; PubMed=6794640;
 RA Metz-Boutigue M.-H., Mazurier J., Jolles J., Spik G., Montreuil J.,
 RA Jolles P.;
 RT "The present state of the human lactotransferrin sequence. Study and
 alignment of the cyanogen bromide fragments and characterization of
 N- and C-terminal domains.";
 RL Biochim. Biophys. Acta 670:243-254(1981).
 RN [11]
 RP SEQUENCE OF 609-711.
 RX MEDLINE=82262043; PubMed=7049727;
 RA Metz-Boutigue M.-H., Jolles J., Mazurier J., Spik G., Montreuil J.,
 RA Jolles P.;
 RT "An 88 amino acid long C-terminal sequence of human
 lactotransferrin.";
 RL FEBS Lett. 142:107-110(1982).
 RN [12]
 RP SEQUENCE OF 436-711 FROM N.A.
 RX MEDLINE=88001031; PubMed=3477300;
 RA Rado T.A., Wei X., Benz E.J. Jr.;
 RT "Isolation of lactoferrin cDNA from a human myeloid library and
 expression of mRNA during normal and leukemic myelopoiesis.";
 RL Blood 70:989-993(1987).
 RN [13]
 RP SEQUENCE OF 237-711 FROM N.A.
 RA McCombie W.R., Wilson R., Chen E., Gibbs R., Zuo L., Johnson D.,
 RA Nhan M., Parnell L., Dedhia N., Ansari A., Mardis E., Schutz K.,
 RA Gnoj L., la Bastide M., Kaplan N., Greco T., Touchman J., Muzny D.,
 RA Chen C.N., Evans C., Fitzgerald M., See L.H., Tang M., Porcel B.M.,
 RA Dragan Y., Giacalone J., Pae A., Powell E., Solinsky K.A., Desilva U.,
 RA Diaz-Perez S., Zhou X., Yu Y., Watanabe M., Doggett N., Garcia D.,
 RA Sagripanti J.L.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [14]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND REVISIONS.
 RX MEDLINE=90064528; PubMed=2585506;
 RA Anderson B.F., Baker H.M., Norris G.E., Rice D.W., Baker E.N.;
 RT "Structure of human lactoferrin: crystallographic structure analysis
 and refinement at 2.8-A resolution.";

RESULT 10	TRFL CAPHI
ID	TRFL
AC	Q2947
DT	15-DEE
DT	15-DEE
DT	10-OC
DE	Lacto
GN	LTF.
OS	Capra
OC	Eukar
OC	Mamma
OC	Bovid
OX	NCBI
RN	[1

SEQUENCE FROM N.A.
 RC TISSUE-Mammary gland;
 RA Lee T., Yu S., Kim S., Lee K., Yu D.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Mammary gland;
 RX MEDLINE=94380047; PubMed=8093048;
 RA le Provost F., Nocart M., Guerin G., Martin P.;
 RT "Characterization of the goat lactoferrin cDNA. Assignment of the
 relevant locus to bovine U12 syntenic group.";
 RL Biochem. Biophys. Res. Commun. 203:1324-1332(1994).
 CC -!- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 OF AN ANION, USUALLY BICARBONATE.
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- DOMAIN: Composed of two homologous domains.
 CC -!- SIMILARITY: Belongs to the transferrin family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; U53857; AAA97958.1; -;
 DR EMBL; X78902; CAA55517.1; -;
 DR HSSP; 077698; 1C82.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN 1; 2.
 DR PROSITE; PS00206; TRANSFERRIN 2; 2.
 DR PROSITE; PS00207; TRANSFERRIN 3; 2.
 KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 Signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 708 LACTOTRANSFERRIN.
 FT REPEAT 20 363 1.
 FT REPEAT 364 708 2.
 FT DISULFID 28 64 BY SIMILARITY.
 FT DISULFID 38 55 BY SIMILARITY.
 FT DISULFID 134 217 BY SIMILARITY.
 FT DISULFID 176 192 BY SIMILARITY.
 FT DISULFID 189 200 BY SIMILARITY.
 FT DISULFID 250 264 BY SIMILARITY.
 FT DISULFID 367 399 BY SIMILARITY.
 FT DISULFID 377 390 BY SIMILARITY.
 FT DISULFID 424 703 BY SIMILARITY.
 FT DISULFID 444 666 BY SIMILARITY.
 FT DISULFID 476 551 BY SIMILARITY.
 FT DISULFID 500 694 BY SIMILARITY.
 FT DISULFID 510 524 BY SIMILARITY.
 FT DISULFID 521 534 BY SIMILARITY.
 FT DISULFID 592 606 BY SIMILARITY.
 FT DISULFID 644 649 BY SIMILARITY.
 FT METAL 79 79 IRON 1 (BY SIMILARITY).
 FT METAL 111 111 IRON 1 (BY SIMILARITY).
 FT METAL 211 211 IRON 1 (BY SIMILARITY).
 FT METAL 212 212 IRON 1 (BY SIMILARITY).
 FT METAL 414 414 IRON 2 (BY SIMILARITY).
 FT METAL 452 452 IRON 2 (BY SIMILARITY).
 FT METAL 545 545 IRON 2 (BY SIMILARITY).
 FT METAL 614 614 IRON 2 (BY SIMILARITY).
 FT BINDING 140 140 CARBONATE 1 (BY SIMILARITY).
 FT BINDING 142 142 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT BINDING 143 143 CARBONATE 1 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).

FT BINDING 478 478 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 482 482 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 484 484 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT BINDING 485 485 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 56 56 I -> V (IN REF. 2).
 FT CONFLICT 88 88 L -> R (IN REF. 2).
 FT CONFLICT 124 124 Q -> K (IN REF. 2).
 FT CONFLICT 154 154 F -> P (IN REF. 2).
 FT CONFLICT 304 304 S -> R (IN REF. 2).
 FT CONFLICT 414 414 D -> G (IN REF. 2).
 SQ SEQUENCE 708 AA; 77358 MW; F2EDA3C83539960D CRC64;
 Query Match 32.0%; Score 1258; DB 1; Length 708;
 Best Local Similarity 40.8%; Pred. No. 3.9e-88;
 Matches 305; Conservative 110; Mismatches 243; Indels 90; Gaps 30;
 QY 6 GALMLLLALRTVLGGMEVRWCATSDPEQHKCCNMSEAFREAGIQPSLLCVRGTSADHCVQ 65
 DB 12 GALGLCLAAPR-----KNVRWCAISLPWSKCYQWRRMRKLG-APSIICIRTSALECIR 66
 QY 66 LIAQEAQDAITLDGAIYEAGKE-HGLKPVVGEVVDQEVG--TSYYAVAVVRSSHVITD 122
 DB 67 AIAGKNADAVTLDSGMVFENGLDPYKLRPVAAEIYGTESKSPQTHYAYAVVKKGSNFOLD 126
 QY 123 TLKGVKSCHTGINRTVGNVVPVGYL-----VESGRLSVMGCDVLKAVSDYFGGSCVPGA 176
 DB 127 QLQGQKSCMGLGRSAGWNI PVGILRPFLSWTESAE-----PLOGAVARFFSASCVPCV 180
 QY 177 GETSYSESLRCRGDSSGEGVCDKSPLERYDYSGAFRCLEAGDGVAFVHKSTVLENT 236
 DB 181 DGKAY-PNLQCLCKG--VGENKCASSQEPYFGYSYGFCLQDQAGDGVAFVKKETTVFEN- 236
 QY 237 DKTLPSPWGQALLSQDFELLCRDGRADVETWRQCHLARVPAHVVVRA-DTDGGILFRL 295
 DB 237 ----LPERAD--RDQYELLCLNNTAPVDAPAFKECHLAQVPSHAVVARSVDGKENVL 289
 QY 296 LNEGQRLP-SHEGSSQFMSSEAVGQKDLLFKDSTSELVPIATOTYEA-WLGEHYLHAMK 353
 DB 290 LRKAQKEGKNKSQSFQLFGSP-GRDRLFPKDSALGFVRIPSKVDSALYLSRILTALK 348
 QY 354 GILCDPNRPL---PYLRMCVLSTPEIQKCGDMAVAFRRQRRLKPEIQCVSAKSPQHCMERI 410
 DB 349 NLRETAELKARCTRVVWCAVGPESQKQOWS-----EQSQNVTCATASTDDCIALV 403
 QY 411 QAEQVDVTLGEDIYTAGKYGILVPAAGEH-----YAPED-----SSNSYVVAVVRDS 461
 DB 404 LKGEADALSLDGGYIYTAG-KCGLVPVMAENRKSSKSYSLDCVLRFTTEGYLAVAVVKK-A 461
 QY 462 SHAFITDELGRKRSCHAGFGSPAGWDVPVGGALI-ORGFIRPKCDVLTAVSFFENASCVP 520
 DB 462 NEGLTWNLSLKGKKSCHTAVDRTAGWNI PMGLTANQTG-----SC-----AFDEFFSQSCAP 512
 QY 521 VNNPKNYPSSLCALCVGDEQGRNCKVGNQRYGYRGAFRCLEVENAGDVAFRVHTTVPD 580
 DB 513 GADPK---SSLCALCAGDQGLDKCVPSNKEKYGYTGAFRCLEADVEDGVAFVKNDTWE 569
 QY 581 NTNGHNSFPWAAERSEDEYELLCNPGARAEVSQFAACNLAIQIPPHAVMVRPDTNITFTVYG 640
 DB 570 NTNGESSADWAKNLNRDFRLCLDGTGTPVTEAQSCLYLAAPNHAVVRSDRAAHEQV 629
 QY 641 LLDKAAQDLFGDDHNKNG-----FKWFDSSNVHGGDILLFKDQATVRAVPVGEKTTVGMWG 694
 DB 630 LLHQ-QALFG-----KNGKNCPDQPCLFKSET---KNLLFNQNTCLAKLGRPTKEYLIG 681
 QY 695 LDYVAALGEMSSQCSGAAAPAPGAPLL 722
 FT BINDING 478 478 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 482 482 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 484 484 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT BINDING 485 485 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 56 56 I -> V (IN REF. 2).
 FT CONFLICT 88 88 L -> R (IN REF. 2).
 FT CONFLICT 124 124 Q -> K (IN REF. 2).
 FT CONFLICT 154 154 F -> P (IN REF. 2).
 FT CONFLICT 304 304 S -> R (IN REF. 2).
 FT CONFLICT 414 414 D -> G (IN REF. 2).
 SQ SEQUENCE 708 AA; 77358 MW; F2EDA3C83539960D CRC64;
 Query Match 32.0%; Score 1258; DB 1; Length 708;
 Best Local Similarity 40.8%; Pred. No. 3.9e-88;
 Matches 305; Conservative 110; Mismatches 243; Indels 90; Gaps 30;
 QY 6 GALMLLLALRTVLGGMEVRWCATSDPEQHKCCNMSEAFREAGIQPSLLCVRGTSADHCVQ 65
 DB 12 GALGLCLAAPR-----KNVRWCAISLPWSKCYQWRRMRKLG-APSIICIRTSALECIR 66
 QY 66 LIAQEAQDAITLDGAIYEAGKE-HGLKPVVGEVVDQEVG--TSYYAVAVVRSSHVITD 122
 DB 67 AIAGKNADAVTLDSGMVFENGLDPYKLRPVAAEIYGTESKSPQTHYAYAVVKKGSNFOLD 126
 QY 123 TLKGVKSCHTGINRTVGNVVPVGYL-----VESGRLSVMGCDVLKAVSDYFGGSCVPGA 176
 DB 127 QLQGQKSCMGLGRSAGWNI PVGILRPFLSWTESAE-----PLOGAVARFFSASCVPCV 180
 QY 177 GETSYSESLRCRGDSSGEGVCDKSPLERYDYSGAFRCLEAGDGVAFVHKSTVLENT 236
 DB 181 DGKAY-PNLQCLCKG--VGENKCASSQEPYFGYSYGFCLQDQAGDGVAFVKKETTVFEN- 236
 QY 237 DKTLPSPWGQALLSQDFELLCRDGRADVETWRQCHLARVPAHVVVRA-DTDGGILFRL 295
 DB 237 ----LPERAD--RDQYELLCLNNTAPVDAPAFKECHLAQVPSHAVVARSVDGKENVL 289
 QY 296 LNEGQRLP-SHEGSSQFMSSEAVGQKDLLFKDSTSELVPIATOTYEA-WLGEHYLHAMK 353
 DB 290 LRKAQKEGKNKSQSFQLFGSP-GRDRLFPKDSALGFVRIPSKVDSALYLSRILTALK 348
 QY 354 GILCDPNRPL---PYLRMCVLSTPEIQKCGDMAVAFRRQRRLKPEIQCVSAKSPQHCMERI 410
 DB 349 NLRETAELKARCTRVVWCAVGPESQKQOWS-----EQSQNVTCATASTDDCIALV 403
 QY 411 QAEQVDVTLGEDIYTAGKYGILVPAAGEH-----YAPED-----SSNSYVVAVVRDS 461
 DB 404 LKGEADALSLDGGYIYTAG-KCGLVPVMAENRKSSKSYSLDCVLRFTTEGYLAVAVVKK-A 461
 QY 462 SHAFITDELGRKRSCHAGFGSPAGWDVPVGGALI-ORGFIRPKCDVLTAVSFFENASCVP 520
 DB 462 NEGLTWNLSLKGKKSCHTAVDRTAGWNI PMGLTANQTG-----SC-----AFDEFFSQSCAP 512
 QY 521 VNNPKNYPSSLCALCVGDEQGRNCKVGNQRYGYRGAFRCLEVENAGDVAFRVHTTVPD 580
 DB 513 GADPK---SSLCALCAGDQGLDKCVPSNKEKYGYTGAFRCLEADVEDGVAFVKNDTWE 569
 QY 581 NTNGHNSFPWAAERSEDEYELLCNPGARAEVSQFAACNLAIQIPPHAVMVRPDTNITFTVYG 640
 DB 570 NTNGESSADWAKNLNRDFRLCLDGTGTPVTEAQSCLYLAAPNHAVVRSDRAAHEQV 629
 QY 641 LLDKAAQDLFGDDHNKNG-----FKWFDSSNVHGGDILLFKDQATVRAVPVGEKTTVGMWG 694
 DB 630 LLHQ-QALFG-----KNGKNCPDQPCLFKSET---KNLLFNQNTCLAKLGRPTKEYLIG 681
 QY 695 LDYVAALGEMSSQCSGAAAPAPGAPLL 722
 FT BINDING 478 478 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 482 482 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 484 484 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT BINDING 485 485 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 56 56 I -> V (IN REF. 2).
 FT CONFLICT 88 88 L -> R (IN REF. 2).
 FT CONFLICT 124 124 Q -> K (IN REF. 2).
 FT CONFLICT 154 154 F -> P (IN REF. 2).
 FT CONFLICT 304 304 S -> R (IN REF. 2).
 FT CONFLICT 414 414 D -> G (IN REF. 2).
 SQ SEQUENCE 708 AA; 77358 MW; F2EDA3C83539960D CRC64;
 Query Match 32.0%; Score 1258; DB 1; Length 708;
 Best Local Similarity 40.8%; Pred. No. 3.9e-88;
 Matches 305; Conservative 110; Mismatches 243; Indels 90; Gaps 30;
 QY 6 GALMLLLALRTVLGGMEVRWCATSDPEQHKCCNMSEAFREAGIQPSLLCVRGTSADHCVQ 65
 DB 12 GALGLCLAAPR-----KNVRWCAISLPWSKCYQWRRMRKLG-APSIICIRTSALECIR 66
 QY 66 LIAQEAQDAITLDGAIYEAGKE-HGLKPVVGEVVDQEVG--TSYYAVAVVRSSHVITD 122
 DB 67 AIAGKNADAVTLDSGMVFENGLDPYKLRPVAAEIYGTESKSPQTHYAYAVVKKGSNFOLD 126
 QY 123 TLKGVKSCHTGINRTVGNVVPVGYL-----VESGRLSVMGCDVLKAVSDYFGGSCVPGA 176
 DB 127 QLQGQKSCMGLGRSAGWNI PVGILRPFLSWTESAE-----PLOGAVARFFSASCVPCV 180
 QY 177 GETSYSESLRCRGDSSGEGVCDKSPLERYDYSGAFRCLEAGDGVAFVHKSTVLENT 236
 DB 181 DGKAY-PNLQCLCKG--VGENKCASSQEPYFGYSYGFCLQDQAGDGVAFVKKETTVFEN- 236
 QY 237 DKTLPSPWGQALLSQDFELLCRDGRADVETWRQCHLARVPAHVVVRA-DTDGGILFRL 295
 DB 237 ----LPERAD--RDQYELLCLNNTAPVDAPAFKECHLAQVPSHAVVARSVDGKENVL 289
 QY 296 LNEGQRLP-SHEGSSQFMSSEAVGQKDLLFKDSTSELVPIATOTYEA-WLGEHYLHAMK 353
 DB 290 LRKAQKEGKNKSQSFQLFGSP-GRDRLFPKDSALGFVRIPSKVDSALYLSRILTALK 348
 QY 354 GILCDPNRPL---PYLRMCVLSTPEIQKCGDMAVAFRRQRRLKPEIQCVSAKSPQHCMERI 410
 DB 349 NLRETAELKARCTRVVWCAVGPESQKQOWS-----EQSQNVTCATASTDDCIALV 403
 QY 411 QAEQVDVTLGEDIYTAGKYGILVPAAGEH-----YAPED-----SSNSYVVAVVRDS 461
 DB 404 LKGEADALSLDGGYIYTAG-KCGLVPVMAENRKSSKSYSLDCVLRFTTEGYLAVAVVKK-A 461
 QY 462 SHAFITDELGRKRSCHAGFGSPAGWDVPVGGALI-ORGFIRPKCDVLTAVSFFENASCVP 520
 DB 462 NEGLTWNLSLKGKKSCHTAVDRTAGWNI PMGLTANQTG-----SC-----AFDEFFSQSCAP 512
 QY 521 VNNPKNYPSSLCALCVGDEQGRNCKVGNQRYGYRGAFRCLEVENAGDVAFRVHTTVPD 580
 DB 513 GADPK---SSLCALCAGDQGLDKCVPSNKEKYGYTGAFRCLEADVEDGVAFVKNDTWE 569
 QY 581 NTNGHNSFPWAAERSEDEYELLCNPGARAEVSQFAACNLAIQIPPHAVMVRPDTNITFTVYG 640
 DB 570 NTNGESSADWAKNLNRDFRLCLDGTGTPVTEAQSCLYLAAPNHAVVRSDRAAHEQV 629
 QY 641 LLDKAAQDLFGDDHNKNG-----FKWFDSSNVHGGDILLFKDQATVRAVPVGEKTTVGMWG 694
 DB 630 LLHQ-QALFG-----KNGKNCPDQPCLFKSET---KNLLFNQNTCLAKLGRPTKEYLIG 681
 QY 695 LDYVAALGEMSSQCSGAAAPAPGAPLL 722
 FT BINDING 478 478 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 482 482 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 484 484 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT BINDING 485 485 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 56 56 I -> V (IN REF. 2).
 FT CONFLICT 88 88 L -> R (IN REF. 2).
 FT CONFLICT 124 124 Q -> K (IN REF. 2).
 FT CONFLICT 154 154 F -> P (IN REF. 2).
 FT CONFLICT 304 304 S -> R (IN REF. 2).
 FT CONFLICT 414 414 D -> G (IN REF. 2).
 SQ SEQUENCE 708 AA; 77358 MW; F2EDA3C83539960D CRC64;
 Query Match 32.0%; Score 1258; DB 1; Length 708;
 Best Local Similarity 40.8%; Pred. No. 3.9e-88;
 Matches 305; Conservative 110; Mismatches 243; Indels 90; Gaps 30;
 QY 6 GALMLLLALRTVLGGMEVRWCATSDPEQHKCCNMSEAFREAGIQPSLLCVRGTSADHCVQ 65
 DB 12 GALGLCLAAPR-----KNVRWCAISLPWSKCYQWRRMRKLG-APSIICIRTSALECIR 66
 QY 66 LIAQEAQDAITLDGAIYEAGKE-HGLKPVVGEVVDQEVG--TSYYAVAVVRSSHVITD 122
 DB 67 AIAGKNADAVTLDSGMVFENGLDPYKLRPVAAEIYGTESKSPQTHYAYAVVKKGSNFOLD 126
 QY 123 TLKGVKSCHTGINRTVGNVVPVGYL-----VESGRLSVMGCDVLKAVSDYFGGSCVPGA 176
 DB 127 QLQGQKSCMGLGRSAGWNI PVGILRPFLSWTESAE-----PLOGAVARFFSASCVPCV 180
 QY 177 GETSYSESLRCRGDSSGEGVCDKSPLERYDYSGAFRCLEAGDGVAFVHKSTVLENT 236
 DB 181 DGKAY-PNLQCLCKG--VGENKCASSQEPYFGYSYGFCLQDQAGDGVAFVKKETTVFEN- 236
 QY 237 DKTLPSPWGQALLSQDFELLCRDGRADVETWRQCHLARVPAHVVVRA-DTDGGILFRL 295
 DB 237 ----LPERAD--RDQYELLCLNNTAPVDAPAFKECHLAQVPSHAVVARSVDGKENVL 289
 QY 296 LNEGQRLP-SHEGSSQFMSSEAVGQKDLLFKDSTSELVPIATOTYEA-WLGEHYLHAMK 353
 DB 290 LRKAQKEGKNKSQSFQLFGSP-GRDRLFPKDSALGFVRIPSKVDSALYLSRILTALK 348
 QY 354 GILCDPNRPL---PYLRMCVLSTPEIQKCGDMAVAFRRQRRLKPEIQCVSAKSPQHCMERI 410
 DB 349 NLRETAELKARCTRVVWCAVGPESQKQOWS-----EQSQNVTCATASTDDCIALV 403
 QY 411 QAEQVDVTLGEDIYTAGKYGILVPAAGEH-----YAPED-----SSNSYVVAVVRDS 461
 DB 404 LKGEADALSLDGGYIYTAG-KCGLVPVMAENRKSSKSYSLDCVLRFTTEGYLAVAVVKK-A 461
 QY 462 SHAFITDELGRKRSCHAGFGSPAGWDVPVGGALI-ORGFIRPKCDVLTAVSFFENASCVP 520
 DB 462 NEGLTWNLSLKGKKSCHTAVDRTAGWNI PMGLTANQTG-----SC-----AFDEFFSQSCAP 512
 QY 521 VNNPKNYPSSLCALCVGDEQGRNCKVGNQRYGYRGAFRCLEVENAGDVAFRVHTTVPD 580
 DB 513 GADPK---SSLCALCAGDQGLDKCVPSNKEKYGYTGAFRCLEADVEDGVAFVKNDTWE 569
 QY 581 NTNGHNSFPWAAERSEDEYELLCNPGARAEVSQFAACNLAIQIPPHAVMVRPDTNITFTVYG 640
 DB 570 NTNGESSADWAKNLNRDFRLCLDGTGTPVTEAQSCLYLAAPNHAVVRSDRAAHEQV 629
 QY 641 LLDKAAQDLFGDDHNKNG-----FKWFDSSNVHGGDILLFKDQATVRAVPVGEKTTVGMWG 694
 DB 630 LLHQ-QALFG-----KNGKNCPDQPCLFKSET---KNLLFNQNTCLAKLGRPTKEYLIG 681
 QY 695 LDYVAALGEMSSQCSGAAAPAPGAPLL 722
 FT BINDING 478 478 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 482 482 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 484 484 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT BINDING 485 485 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 56 56 I -> V (IN REF. 2).
 FT CONFLICT 88 88 L -> R (IN REF. 2).
 FT CONFLICT 124 124 Q -> K (IN REF. 2).
 FT CONFLICT 154 154 F -> P (IN REF. 2).
 FT CONFLICT 304 304 S -> R (IN REF. 2).
 FT CONFLICT 414 414 D -> G (IN REF. 2).
 SQ SEQUENCE 708 AA; 77358 MW; F2EDA3C83539960D CRC64;
 Query Match 32.0%; Score 1258; DB 1; Length 708;
 Best Local Similarity 40.8%; Pred. No. 3.9e-88;
 Matches 305; Conservative 110; Mismatches 243; Indels 90; Gaps 30;
 QY 6 GALMLLLALRTVLGGMEVRWCATSDPEQHKCCNMSEAFREAGIQPSLLCVRGTSADHCVQ 65
 DB 12 GALGLCLAAPR-----KNVRWCAISLPWSKCYQWRRMRKLG-APSIICIRTSALECIR 66
 QY 66 LIAQEAQDAITLDGAIYEAGKE-HGLKPVVGEVVDQEVG--TSYYAVAVVRSSHVITD 122
 DB 67 AIAGKNADAVTLDSGMVFENGLDPYKLRPVAAEIYGTESKSPQTHYAYAVVKKGSNFOLD 126
 QY 123 TLKGVKSCHTGINRTVGNVVPVGYL-----VESGRLSVMGCDVLKAVSDYFGGSCVPGA 176
 DB 127 QLQGQKSCMGLGRSAGWNI PVGILRPFLSWTESAE-----PLOGAVARFFSASCVPCV 180
 QY 177 GETSYSESLRCRGDSSGEGVCDKSPLERYDYSGAFRCLEAGDGVAFVHKSTVLENT 236
 DB 181 DGKAY-PNLQCLCKG--VGENKCASSQEPYFGYSYGFCLQDQAGDGVAFVKKETTVFEN- 236
 QY 237 DKTLPSPWGQALLSQDFELLCRDGRADVETWRQCHLARVPAHVVVRA-DTDGGILFRL 295
 DB 237 ----LPERAD--RDQYELLCLNNTAPVDAPAFKECHLAQVPSHAVVARSVDGKENVL 289
 QY 296 LNEGQRLP-SHEGSSQFMSSEAVGQKDLLFKDSTSELVPIATOTYEA-WLGEHYLHAMK 353
 DB 290 LRKAQKEGKNKSQSFQLFGSP-GRDRLFPKDSALGFVRIPSKVDSALYLSRILTALK 348
 QY 354 GILCDPNRPL---PYLRMCVLSTPEIQKCGDMAVAFRRQRRLKPEIQCVSAKSPQHCMERI 410
 DB 349 NLRETAELKARCTRVVWCAVGPESQKQOWS-----EQSQNVTCATASTDDCIALV 403
 QY 411 QAEQVDVTLGEDIYTAGKYGILVPAAGEH-----YAPED-----SSNSYVVAVVRDS 461
 DB 404 LKGEADALSLDGGYIYTAG-KCGLVPVMAENRKSSKSYSLDCVLRFTTEGYLAVAVVKK-A 461
 QY 462 SHAFITDELGRKRSCHAGFGSPAGWDVPVGGALI-ORGFIRPKCDVLTAVSFFENASCVP 520
 DB 462 NEGLTWNLSLKGKKSCHTAVDRTAGWNI PMGLTANQTG-----SC-----AFDEFFSQSCAP 512
 QY 521 VNNPKNYPSSLCALCVGDEQGRNCKVGNQRYGYRGAFRCLEVENAGDVAFRVHTTVPD 580
 DB 513 GADPK---SSLCALCAGDQGLDKCVPSNKEKYGYTGAFRCLEADVEDGVAFVKNDTWE 569
 QY 581 NTNGHNSFPWAAERSEDEYELLCNPGARAEVSQFAACNLAIQIPPHAVMVRPDTNITFTVYG 640
 DB 570 NTNGESSADWAKNLNRDFRLCLDGTGTPVTEAQSCLYLAAPNHAVVRSDRAAHEQV 629
 QY 641 LLDKAAQDLFGDDHNKNG-----FKWFDSSNVHGGDILLFKDQATVRAVPVGEKTTVGMWG 694
 DB 630 LLHQ-QALFG-----KNGKNCPDQPCLFKSET---KNLLFNQNTCLAKLGRPTKEYLIG 681
 QY 695 LDYVAALGEMSSQCSGAAAPAPGAPLL 722
 FT BINDING 478 478 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 482 482 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 484 484 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT BINDING 485 485 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 56 56 I -> V (IN REF. 2).
 FT CONFLICT 88 88 L -> R (IN REF. 2).
 FT CONFLICT 124 124 Q -> K (IN REF. 2).
 FT CONFLICT 154 154 F -> P (IN REF. 2).
 FT CONFLICT 304 304 S -> R (IN REF. 2).
 FT CONFLICT 414 414 D -> G (IN REF. 2).
 SQ SEQUENCE 708 AA; 77358 MW; F2EDA3C83539960D CRC64;
 Query Match 32.0%; Score 1258; DB 1; Length 708;
 Best Local Similarity 40.8%; Pred. No. 3.9e-88;
 Matches 305; Conservative 110; Mismatches 243; Indels 90; Gaps 30;
 QY 6 GALMLLLALRTVLGGMEVRWCATSDPEQHKCCNMSEAFREAGIQPSLLCVRGTSADHCVQ 65
 DB 12 GALGLCLAAPR-----KNVRWCAISLPWSKCYQWRRMRKLG-APSIICIRTSALECIR 66
 QY 66 LIAQEAQDAITLDGAIYEAGKE-HGLKPVVGEVVDQEVG--TSYYAVAVVRSSHVITD 122
 DB 67 AIAGKNADAVTLDSGMVFENGLDPYKLRPVAAEIYGTESKSPQTHYAYAVVKKGSNFOLD 126
 QY 123 TLKGVKSCHTGINRTVGNVVPVGYL-----VESGRLSVMGCDVLKAVSDYFGGSCVPGA 176
 DB 127 QLQGQKSCMGLGRSAGWNI PVGILRPFLSWTESAE-----PLOGAVARFFSASCVPCV 180
 QY 177 GETSYSESLRCRGDSSGEGVCDKSPLERYDYSGAFRCLEAGDGVAFVHKSTVLENT 236
 DB 181 DGKAY-PNLQCLCKG--VGENKCASSQEPYFGYSYGFCLQDQAGDGVAFVKKETTVFEN- 236
 QY 237 DKTLPSPWGQALLSQDFELLCRDGRADVETWRQCHLARVPAHVVVRA-DTDGGILFRL 295
 DB 237 ----LPERAD--RDQYELLCLNNTAPVDAPAFKECHLAQVPSHAVVARSVDGKENVL 289
 QY 296 LNEGQRLP-SHEGSSQFMSSEAVGQKDLLFKDSTSELVPIATOTYEA-WLGEHYLHAMK 353
 DB 290 LRKAQKEGKNKSQSFQLFGSP-GRDRLFPKDSALGFVRIPSKVDSALYLSRILTALK 348
 QY 354 GILCDPNRPL---PYLRMCVLSTPEIQKCGDMAVAFRRQRRLKPEIQCVSAKSPQHCMERI 410
 DB 349 NLRETAELKARCTRVVWCAVGPESQKQOWS-----EQSQNVTCATASTDDCIALV 403
 QY 411 QAEQVDVTLGEDIYTAGKYGILVPAAGEH-----YAPED-----SSNSYVVAVVRDS 461
 DB 404 LKGEADALSLDGGYIYTAG-KCGLVPVMAENRKSSKSYSLDCVLRFTTEGYLAVAVVKK-A 461
 QY 462 SHAFITDELGRKRSCHAGFGSPAGWDVPVGGALI-ORGFIRPKCDVLTAVSFFENASCVP 520
 DB 462 NEGLTWNLSLKGKKSCHTAVDRTAGWNI PMGLTANQTG-----SC-----AFDEFFSQSCAP 512
 QY 521 VNNPKNYPSSLCALCVGDEQGRNCKVGNQRYGYRGAFRCLEVENAGDVAFRVHTTVPD 580
 DB 513 GADPK---SSLCALCAGDQGLDKCVPSNKEKYGYTGAFRCLEADVEDGVAFVKNDTWE 569
 QY 581 NTNGHNSFPWAAERSEDEYELLCNPGARAEVSQFAACNLAIQIPPHAVMVRPDTNITFTVYG 640
 DB 570 NTNGESSADWAKNLNRDFRLCLDGTGTPVTEAQSCLYLAAPNHAVVRSDRAAHEQV 629
 QY 641 LLDKAAQDLFGDDHNKNG-----FKWFDSSNVHGGDILLFKDQATVRAVPVGEKTTVGMWG 694
 DB 630 LLHQ-QALFG-----KNGKNCPDQPCLFKSET---KNLLFNQNTCLAKLGRPTKEYLIG 681
 QY 695 LDYVAALGEMSSQCSGAAAPAPGAPLL 722
 FT BINDING 478 478 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 482 482 CARBONATE 2 (BY SIMILARITY).
 FT BINDING 484 484 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT BINDING 485 485 CARBONATE 2 (VIA AMIDE NITROGEN) (BY
 SIMILARITY).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD

FT	METAL	614	614	IRON 2.	Db	127	QVGRKSGHTGLRSAGWIIIPWGLI-----RPLVSWTESLEPLOGAVAKFFSASCPCIDR	182
FT	BINDING	136	136	CARBONATE 1.	Qy	179	TSYESLCLRCRGSSGEGVCDKSPLERYDYDSYGAFRCLAEAGAGDVAFVXHSIVLENTDG	238
FT	BINDING	140	140	CARBONATE 1.	Db	183	QAY-PNLQCLCKGE--GENQCACSSREPYFGVSGAFKCLQDGAGDVAFVKETTVFEN---	236
FT	BINDING	142	142	CARBONATE 2.	Qy	239	KTLPSWGOALLSQDFELLRCRGSRADVTWEMROCHLARVPAHVVVRA-DTGGGLIFRLIN	297
FT	BINDING	143	143	CARBONATE 2.	Db	237	--LPEKAD--RDQYELLCLNNSRAPVDAFKECHLAQVESHAVVARSVDGKEDLWKILS	291
FT	BINDING	144	144	CARBONATE 2.	Qy	298	EGORLF-SHEGSSFOMESSEAYGOKDLFLKDSLSLPIATQTYEA-NLGHEYHLAMKGL	355
FT	BINDING	145	145	CARBONATE 2.	Db	292	KAQKFGKNGKRSRSPOLFGSP-PGORDLFLKDSLSLPIATQTYEA-NLGHEYHLAMKGL	350
FT	CARBOHYD	146	146	CARBONATE 2.	Qy	356	LCDPRLPP-YLR--WCVLSTPEIQKCGDMVAFRQRQLKPEIQCVSAKSPHQCMERIOA	412
FT	CARBOHYD	147	147	CARBONATE 2.	Db	351	RETAEEVKARVTRVVMCAVGPPEQKCCQWS-----QQSGQNVTCATASTTDDCIVLVLK	405
FT	CARBOHYD	148	148	CARBONATE 2.	Qy	413	EQVDVAVTLSGEDIYTAGKKYGLVPAAGHYHAPEDSSN-----SYVVAVVRRDSSH	463
FT	CARBOHYD	149	149	CARBONATE 2.	Db	406	GEADALNLDGGYIYTAG-KCGLVLPVLAENRKSSKHSLDCVLRPTGYLA VAVVK--ANE	463
FT	CARBOHYD	150	150	CARBONATE 2.	Qy	464	AFTLDLGRKSGCHAGFGSPAGWDVPVGCALI-ORGFIRPKDCDVLTA VSEPFENASCVPVN	522
FT	CARBOHYD	151	151	CARBONATE 2.	Db	464	GLTWNSLADKKSCHTAVDRTAGWNIIPMGLI VNTQT-----SC-----AFDEFFSQSCAPGA	514
FT	CARBOHYD	152	152	CARBONATE 2.	Qy	523	NPQVPPSLCALCVGDEQGRNKCVCNSQERYVYGRGAFRCVLENAGDVAFVRRHTTVFONT	582
FT	CARBOHYD	153	153	CARBONATE 2.	Db	515	DPK---SRLCALCAGDQGLDKVCPNSKEYGYGTGAFRCVLENAGDVAFVRRHTTVFONT	571
FT	CARBOHYD	154	154	CARBONATE 2.	Qy	583	NGHNSPWAELRSEDYELLCPNGARAEVSQFAACNLAQIIPPHAVMVPDNTNIFTVYGLL	642
FT	CARBOHYD	155	155	CARBONATE 2.	Db	572	NGESTADWAKNLNREDFRLCLDGTGRKPVTEAQSCHLA VAPNHAVVSRDRAAHVQKVL	631
FT	CARBOHYD	156	156	CARBONATE 2.	Qy	643	DKAODLFGDDHKNKG-----FKMFDSSNYHGDLLFKDATTAVRVPVGEKTYRGWGLD	696
FT	CARBOHYD	157	157	CARBONATE 2.	Db	632	HQ-QALFG-----KNGKNCPPDKFLFKSET---KNLLFNDNTECLAKUGRPTVEYLGTE	683
FT	CARBOHYD	158	158	CARBONATE 2.	Qy	697	YVAALGMSQQCSGAAAPAPCAPIL 722	
FT	CARBOHYD	159	159	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	160	160	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	161	161	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	162	162	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	163	163	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	164	164	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	165	165	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	166	166	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	167	167	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	168	168	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	169	169	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	170	170	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	171	171	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	172	172	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	173	173	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	174	174	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	175	175	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	176	176	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	177	177	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	178	178	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	179	179	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	180	180	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	181	181	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	182	182	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	183	183	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	184	184	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	185	185	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	186	186	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	187	187	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	188	188	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	189	189	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	190	190	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	191	191	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	192	192	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	193	193	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	194	194	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	195	195	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	196	196	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	197	197	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	198	198	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	199	199	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	200	200	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	201	201	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	202	202	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	203	203	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	204	204	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	205	205	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	206	206	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	207	207	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	208	208	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	209	209	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	210	210	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	211	211	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	212	212	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	213	213	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	214	214	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	215	215	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	216	216	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	217	217	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	218	218	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	219	219	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	220	220	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	221	221	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	222	222	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	223	223	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	224	224	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	225	225	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	226	226	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	227	227	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	228	228	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	229	229	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	230	230	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	231	231	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	232	232	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	233	233	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	234	234	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	235	235	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	236	236	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	237	237	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	238	238	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	239	239	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	240	240	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	241	241	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	242	242	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	243	243	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	244	244	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	245	245	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	246	246	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	247	247	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	248	248	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	249	249	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	250	250	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	251	251	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	252	252	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	253	253	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	254	254	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	255	255	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	256	256	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	257	257	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	258	258	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	259	259	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	260	260	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	261	261	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	262	262	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	263	263	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	264	264	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	265	265	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	266	266	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	267	267	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	268	268	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	269	269	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	270	270	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	271	271	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	272	272	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	273	273	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	274	274	CARBONATE 2.	Qy	684	YVTAIANL--KKCS-----TSPIL 700	
FT	CARBOHYD	275	275	CARBONATE 2.	Db	684	YVTAIANL--KKCS-----TSPIL 700	
FT								


```

RX MEDLINE=91274362; PubMed=2054387;
RA Banfield D.K., Chow B.K.-C., Funk W.D., Robertson K.A., Umelas T.M.,
RA Woodworth R.C., Maggillivray R.T.A.;
RT "The nucleotide sequence of rabbit liver transferrin cDNA.";
RL Biochim. Biophys. Acta 1089:262-265(1991).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=New Zealand white;
RA Ghareeb B.A.A., Thepot D., Puissant C., Cajero-Juarez M.,
RA Houdebine L.M.;
RT "Cloning and structural organisation of the rabbit transferrin encoding
RT gene";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE OF 20-51.
RP MEDLINE=88209278; PubMed=3365331;
RA Godovac-Zimmermann J.;
RT "Isolation, characterization and N-terminal amino-acid sequence of
RT rabbit transferrin.";
RL Biol. Chem. Hoppe-Seyler 369:93-96(1988).
RN [4]
RN SEQUENCE OF 483-545.
RP MEDLINE=89005676; PubMed=3169252;
RA Evans R.W., Aitken A., Patel K.J.;
RT "Evidence for a single glycan moiety in rabbit serum transferrin and
RT location of the glycan within the polypeptide chain.";
RL FEBS Lett. 238:39-42(1988).
RN [5]
RN X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RC TISSUE=Plasma;
RX MEDLINE=89026775; PubMed=3179277;
RA Bailey S., Evans R.W., Garratt R.C., Gorinsky B., Hasnain S.,
RA Horsburgh C., Jhoti H., Lindley P.F., Mydin A., Sarra R., Watson J.L.;
RT "Molecular structure of serum transferrin at 3.3-A resolution.";
RL Biochemistry 27:5804-5812(1988).
RN [6]
RN X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RA Sarra R., Garratt R.C., Gorinsky B., Jhoti H., Lindley P.F.;
RT "High-resolution X-ray studies on rabbit serum transferrin:
RT preliminary structure analysis of the N-terminal half-molecule at
RT 2.3-A resolution.";
RL Acta Crystallogr. B 46:763-771(1990).
CC -1- FUNCTION: Transferrins are iron binding transport proteins which
CC can bind two atoms of ferric iron in association with the binding
CC of an anion, usually bicarbonate. It is responsible for the
CC transport of iron from sites of absorption and heme degradation to
CC those of storage and utilization. Serum transferrin may also have
CC a further role in stimulating cell proliferation.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.
CC -1- DOMAIN: Composed of two homologous domains.
CC -1- SIMILARITY: Belongs to the transferrin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X58533; CAA41424.1; --
DR EMBL; AF031625; AAB94136.1; --
DR EMBL; AF031611; AAB94136.1; JOINED.
DR EMBL; AF031612; AAB94136.1; JOINED.
DR EMBL; AF031613; AAB94136.1; JOINED.
DR EMBL; AF031614; AAB94136.1; JOINED.
DR EMBL; AF031615; AAB94136.1; JOINED.
DR EMBL; AF031616; AAB94136.1; JOINED.
DR EMBL; AF031617; AAB94136.1; JOINED.
DR EMBL; AF031618; AAB94136.1; JOINED.
DR EMBL; AF031619; AAB94136.1; JOINED.
DR EMBL; AF031620; AAB94136.1; JOINED.
DR EMBL; AF031621; AAB94136.1; JOINED.
DR EMBL; AF031622; AAB94136.1; JOINED.
DR EMBL; AF031623; AAB94136.1; JOINED.
DR EMBL; AF031624; AAB94136.1; JOINED.
DR PDB; 1TFD; 15-APR-93.
DR PDB; 1JNF; 09-JAN-02.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00094; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 1.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
KW signal; 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 695 SEROTRANSFERRIN.
FT REPEAT 20 355 1.
FT REPEAT 356 695 2.
FT DISULFID 28 67
FT DISULFID 38 58
FT DISULFID 137 213
FT DISULFID 156 350
FT DISULFID 177 193
FT DISULFID 180 196
FT DISULFID 190 198
FT DISULFID 246 260
FT DISULFID 358 612
FT DISULFID 364 396
FT DISULFID 374 387
FT DISULFID 421 690
FT DISULFID 436 653
FT DISULFID 468 539
FT DISULFID 492 681
FT DISULFID 502 516
FT DISULFID 513 522
FT DISULFID 579 593
FT DISULFID 631 636
FT CARBOHYD 509 509 N-LINKED (GLCNAC. . .).
FT METAL 82 82 IRON 1.
FT METAL 114 114 IRON 1.
FT METAL 207 207 IRON 1.
FT METAL 268 268 IRON 1.
FT METAL 411 411 IRON 2.
FT METAL 444 444 IRON 2.
FT METAL 533 533 IRON 2.
FT METAL 601 601 IRON 2.
FT BINDING 139 139 CARBONATE 1.
FT BINDING 143 143 CARBONATE 1.
FT BINDING 145 145 CARBONATE 1 (VIA AMIDE NITROGEN).
FT BINDING 146 146 CARBONATE 1 (VIA AMIDE NITROGEN).
FT BINDING 470 470 CARBONATE 2.
FT BINDING 474 474 CARBONATE 2.
FT BINDING 476 476 CARBONATE 2 (VIA AMIDE NITROGEN).
FT BINDING 477 477 CARBONATE 2 (VIA AMIDE NITROGEN).
FT VARIANT 517 517 V -> I.
FT CONFLICT 7 7 MISSING (IN REF. 1).
FT CONFLICT 47 47 K -> S (IN REF. 3).
FT CONFLICT 50 50 P -> Y (IN REF. 3).
FT STRAND 27 29
FT HELIX 32 45
FT TURN 46 48
FT TURN 59 60
FT STRAND 64 72
FT TURN 73 74
FT STRAND 78 81
FT HELIX 83 90
FT TURN 91 94
FT STRAND 96 102
FT STRAND 114 121
FT TURN 122 123
FT TURN 128 129

```

```

FT TURN 132 133
FT STRAND 136 137
FT TURN 141 142
FT TURN 144 147
FT TURN 148 154
FT TURN 155 157
FT TURN 165 170
FT TURN 171 172
FT TURN 176 177
FT TURN 179 180
FT TURN 190 190
FT TURN 194 195
FT TURN 207 214
FT TURN 215 217
FT TURN 221 225
FT TURN 226 227
FT TURN 228 232
FT TURN 236 239
FT TURN 240 241
FT TURN 242 246
FT TURN 247 249
FT TURN 250 253
FT TURN 254 255
FT TURN 257 259
FT TURN 263 265
FT TURN 269 273
FT TURN 279 288
FT TURN 289 291
FT TURN 303 304
FT TURN 306 306
FT TURN 307 308
FT TURN 309 309
FT TURN 316 317
FT TURN 321 322
SQ SEQUENCE 695 AA; 76670 MW; DB12F34D87AE9D55 CRC64;

Query Match
Best Local Similarity 31.5%; Score 1240; DB 1; Length 695;
Matches 308; Conservative 108; Mismatches 231; Indels 98; Gaps 34;

QY 1 MRGPGALWLLALRTVLGME--VRWCATSDPEQHKCGN-----MSEAFREAGIQPSLLC 54
DB 1 MRLAAGALLACAALGLCLAVTEKTVRWCAVNDHEASKCANFRDSMKKVLPEDG--PRIIC 58
QY 55 VRGTSADHCVOIAQAQEAADITLDGAIYEAG-KEHGLKPVVGEVYDQVVG--TSYYAVA 111
DB 59 VKKASYLDCIKIAAHEADAVTLDAGLVHEAGLTPNNLKPVVAEFYGSKENPKTFYYAVA 118
QY 112 VYRRSHVITDILKGKVSCHTGINRTVGWNPVGVYLVESGRLSVMGCD-----VLKAV 164
DB 119 LVKGSNFQINELQKKKCHTGLGRSAGWNIFIGLLY-----CDLPEPRKPLEKAV 169
QY 165 SDYFGSCVPGAGETSYSESRLCRDGSSEGVCDKSPLERYDYYSGAFRCLAEAGADV 224
DB 170 ASFFSGSCVPCADGADFPQ-LQLCPG-----CGCSSVQVPYFGYSGAFKCLKDGLGV 221
QY 225 AFVKHSTVLNTDGTKLPSWGQALLSQDFELLCRDGSRADVTWEQCHLARVPAHAVVR 284
DB 222 AFVKQETIFEN-----LPSKDE---RDQYELLCLDNTKRPVDEYEQCHLARVPSHAVAR 273
QY 285 A-DTGGGLIFRLNEQRLFSHEGS-SFQMFSEAYGOKDLLFKDSTSELVPIATQ-TYE 341
DB 274 SVDGKEDLLIWEINQAEHFQDKSGDFOLFSS-PHG-KULLFKDSAYGFFKVPFRMDAN 331
QY 342 AWLGHYHLAMKGL-----LCDPNRLP---PYLRWCVLSTPEIQKCGDMAVAFRRQLKPEI 395
DB 332 LVLGYEYVAVRNLRREGIC-PDPLQDECKAVKWCALSHERLKCDEWSVTSGGL-----I 385
QY 396 QCVSAKSPQHCWERIOAEQVDAVTLSGEDIYTAGKYGLVPAAGHY-----APEDSS 448
DB 386 ECESABETPDCTAKINMGADAMSLDGGVYIYAG-QCGLVPLVAENYESTDCKKAPEE-- 442
QY 449 NSYVVAVVVRDSSHAFTLDELGRKRSCHAGFGSPAGMDVPVGVGALIQRFIRPKDCDVL 508

```

```

DB 443 -GYLSVAVVKK-SNPDINWNNLEGKKSCHTAVDRTAGNIPMGLL-----YNRINHC----- 492
QY 509 AVSEFFNASCVPVNNPKNYPSSLCALCVGDGGRNKKCVGNQERYGYYGAPRCLVENAG 568
DB 493 RFDEFROGCAP-GSQKN--SSSLCCLVGP-----SVCAPNNREGYYGYTGAFCRLVEK-G 544
QY 569 DVAFVYRHTTVPDNTGNHSEPKAAELRSEDYELLCPCNGARAEVSOFAACNLAIQIPPHAVM 628
DB 545 DVAFVKSQTVLQNTGGRNSEPWAKLKEEDFELLCDGTGRKPVSEAHNCHLAKANHAVV 604
QY 629 VRPDNIPTVYGLLD---KAODLFGDHNKNGKFMFDSNHYHGQDLLFKDATTVAVPVGE 685
DB 605 SRKDKAAACVQKLLDLQVEYGNVTADCSK--FCNFSKST---KOLLFRDDTKCLVDLRG 659
QY 686 KTYRGWLGLDYVAALGEMSSQOCS 710
DB 660 KNTYEKYLGADEYKAVSNL--RKCS 682

RESULT 15
TRFE PAROL
ID TRFE PAROL STANDARD; PRT; 685 AA.
AC O93429;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Serotransferrin precursor.
OS Paralicthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Paralicthidae; Paralicthys.
OC NCBI_TaxID=8255;
RN [1]
RP SEQUENCE FROM N.A.
RA Kim Y., Lee J., Hong Y., Hiroo I., Aoki T.;
RT "Molecular cloning and sequence analysis of transferrin cDNA from
RT Japanese flounder Paralicthys olivaceus,";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: Composed of two homologous domains.
CC -!- SIMILARITY: Belongs to the transferrin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D88801; BAA28944.1; -.
DR HSP; P56410; IAOV.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00094; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 1.
KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
FT SIGNAL.
FT CHAIN 1 16 BY SIMILARITY.
FT CHAIN 17 685 SEROTRANSFERRIN.
FT DISULFID 26 48 BY SIMILARITY.
FT DISULFID 125 206 BY SIMILARITY.
FT DISULFID 170 184 BY SIMILARITY.
FT DISULFID 234 248 BY SIMILARITY.

```

```
FT DISULFID 343 379 BY SIMILARITY.
FT DISULFID 353 370 BY SIMILARITY.
FT DISULFID 404 678 BY SIMILARITY.
FT DISULFID 419 639 BY SIMILARITY.
FT DISULFID 451 526 BY SIMILARITY.
FT DISULFID 475 667 BY SIMILARITY.
FT DISULFID 485 499 BY SIMILARITY.
FT DISULFID 496 509 BY SIMILARITY.
FT DISULFID 566 580 BY SIMILARITY.
FT METAL 72 72 IRON 1 (BY SIMILARITY).
FT METAL 102 102 IRON 1 (BY SIMILARITY).
FT METAL 200 200 IRON 1 (BY SIMILARITY).
FT METAL 256 256 IRON 1 (BY SIMILARITY).
FT METAL 394 394 IRON 2 (BY SIMILARITY).
FT METAL 428 428 IRON 2 (BY SIMILARITY).
FT METAL 520 520 IRON 2 (BY SIMILARITY).
FT METAL 588 588 IRON 2 (BY SIMILARITY).
FT BINDING 127 127 CARBONATE 1 (BY SIMILARITY).
FT BINDING 131 131 CARBONATE 1 (BY SIMILARITY).
FT BINDING 133 133 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 134 134 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 453 453 CARBONATE 2 (BY SIMILARITY).
FT BINDING 457 457 CARBONATE 2 (BY SIMILARITY).
FT BINDING 459 459 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT BINDING 460 460 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).
FT CARBOHYD 476 476 N-LINKED (GLCNAC...) (POTENTIAL).
FT SEQUENCE 685 AA; 74654 MW; 5A6F622AD7D5B662 CRC64;

Query Match 31.4%; Score 1236.5; DB 1; Length 685;
Best Local Similarity 38.8%; Pred. No. 1.6e-86;
Matches 277; Conservative 118; Mismatches 242; Indels 77; Gaps 24;

Qy 10 LLALRTVLGME-VRWCATSDPEQHKGNMSEAFREAGIQPSLLCVRGTSADHCVOLIA 68
Db 9 LLGLATIASCIDTVKVCVTSTKENLKTALAAA-----APVFSCVARASITDCLTAIK 62

Qy 69 AQEADATLGGAIYAG-KEHGLKPVVGEVYQEVGTSYVAVVRRSSHVTIDTLKGV 127
Db 63 AGEADATLGGGEIYTAGLDEYKLPHTIAEQYGTSTDTCTCYAVAVAKKNTGFLHQLMGK 122

Qy 128 KSCHTGINRTVGNVPVGYLVESGRSLVMGCVLK---AVSDYFGGSCVPAGAGTSSVES 184
Db 123 KSCHTGVKSGAGNIPITGLLSMDFTKWKSDDKLUEVVGEFFHSSCAPGATD---SAN 179

Qy 185 LCRLCRGDSGEGVCDK-SPLERYDYSGAFRCLEAGAGDAFVYGHSTVLENTDGTLP 243
Db 180 LCKLCIGD-----CSKSSSETPYNYHGAFQCLKDKGKGDVAFVKHLTVPEE----- 226

Qy 244 WQALLSQDELLCRDGSRADVTWEROCHLARPAPAHAVVVRADTD---GGLI FRLNREGOR 301
Db 227 -----KNDYELLCKONTXRPIDQFENCIDLAKVPASHAVVTRKONEELAQFIWQSLSVK- 279

Qy 302 LFSHEGSSFOMFSEAYGOKDLFKDSTSELVPIATQT-YEAWLGHGYLHAMKGL--LCD 358
Db 280 -----NFNLFSTPYGGKLNLFKSTTTLVQLPLNVDRHTMYLGHYLESVKALKIVNI 332

Qy 359 PNRLLPYLRWCVLSTPEIQCGDMAVAFRRQRKPEIQCVSAKSPQHCHMERIOAEQVDAV 418
Db 333 PSTTSDAMKCAVGRSSDKDSWSVASLVQD-GTTIDCIKGNVTDDCLKKIMHKEADAM 391

Qy 419 TLSGEDIYTAGKYGVLPAAGEHY-----APEDSSNSYVAVVVRDSSHAFTLDELRG 472
Db 392 AVDGGQYITAG-KCGLVPAMVEQYDQGCSCAP-GAARLYYAVAVIKKS--GVTWENLRN 447

Qy 473 KRSCHAGFGSPAGWDVPVVGALIQRGFIRPKDCDLVLTAVSEFFNASCVPVNNPKNYPSSLC 532
Db 448 KRSCHTIGIRNAGWNTPMGLIYE-----QTKNCN-----FSAFFSSSCAPGADPS---SQLC 496

Qy 533 ALCVGDQGRNKCVCNQSERYGYRGAFRCLEVENAGDVAFVRHTTVFDNTNGHNSBPMAA 592
```

```
Db 497 AQCAGNAESINKCKASNEERYAYAGAFRCLEAGKGDVAFVKHISIVKENTDGGQPE-WAK 555
Qy 593 ELRSEDEYELLCPNGARAEVSOFAACNLAQIPPHAVVVRPD--TNIFTVYGLLQKQDLFG 650
Db 556 AFLSNDYELICPSKGPVSVENFMSCNLAKVNAHAVVTRPEIRTKVVT---FLNNQOSHFG 612
Qy 651 DDHNKNGFKMFDSNNYHGQDLFLFKDATVRAVPVGEKTTYRGWLGLDYVAALEGM 704
Db 613 NSASEESFKMFTSPD--GENLLFKYSTKCLQEI PAHLDYKGLGQEVMTVMSSL 664
```

Search completed: May 14, 2004, 09:42:20
Job time : 20 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 14, 2004, 09:40:37 ; Search time 23 Seconds
(without alignments)
1656.520 Million cell updates/sec

Title: US-10-049-957-4
Perfect score: 3936
Sequence: 1 MRGPSALWLLALRTVLGG.....APLPLLLPALAARLLPPAL 738

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3933	99.9	738	6	5262177-2
2	3840	97.6	719	2	US-08-520-933-3
3	3840	97.6	719	4	US-09-285-040-3
4	3792	96.3	717	6	5262177-5
5	1850	47.0	502	2	US-08-459-818-19
6	1850	47.0	502	2	US-08-889-666-19
7	1850	47.0	502	2	US-08-465-078-19
8	1850	47.0	502	2	US-08-725-776-19
9	1850	47.0	502	2	US-08-488-062-19
10	1269.5	32.3	698	2	US-08-175-158A-2
11	1268.5	32.2	1074	2	US-08-470-058-2
12	1268.5	32.2	1074	3	US-09-037-188-2
13	1268.5	32.2	1074	3	US-09-285-310-2
14	1263.5	32.1	1410	2	US-08-470-058-4
15	1263.5	32.1	1410	3	US-09-037-188-4
16	1263.5	32.1	1410	3	US-09-285-310-4
17	1261	32.0	705	2	US-08-655-640-2
18	1260.5	32.0	708	2	US-08-555-640-4
19	1255.5	31.9	711	1	US-08-445-681-2
20	1255.5	31.9	711	1	US-08-250-308-2
21	1255.5	31.9	711	1	US-08-453-703-2
22	1255.5	31.9	711	2	US-08-456-106-2
23	1255.5	31.9	711	3	US-08-456-108-2
24	1255.5	31.9	711	3	US-09-265-577-2
25	1255.5	31.9	711	4	US-09-633-739-2
26	1255.5	31.9	711	5	PCT-US93-03614-2
27	1252	31.8	694	3	US-08-724-586-2

ALIGNMENTS

RESULT 1
5262177-2

; Patent No. 5262177
; APPLICANT: BROWN, J OSEPH P.; ESTIN, CHARLES D.; PLOWMAN, GREGORY
; D.; HELSTROM, KARL E.; ROSE, TIMOTHY M.; HELSTROM, INGERGERD;
; PURCHIO, ANTHONY F.; HU, SHIU-LOK; PENNATHUR, SRIDHAR
; TITLE OF INVENTION: RECOMBINANT VIRUSES ENCODING THE HUMAN
; MELANOMA-ASSOCIATED ANTIGEN
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/7230
; FILING DATE: 27-JAN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 827,313
; FILING DATE: 07-FEB-1986
; SEQ ID NO: 2:
; LENGTH: 738
5262177-2

Query Match 99.9%; Score 3933; DB 6; Length 738;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 737; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MRGPSALWLLALRTVLGGMEVRCATSDPEQHKCGNMSEAPREAGIQBSLLCVRCTSA	60
Db	1	MRGPSALWLLALRTVLGGMEVRCATSDPEQHKCGNMSEAPREAGIQBSLLCVRGTSA	60
Qy	61	DHCVQLIAAQEADAITLDGGAIEAGKEHGLKPVGVEVDQEVGTSYAVAVVRRSSHVT	120
Db	61	DHCVQLIAAQEADAITLDGGAIEAGKEHGLKPVGVEVDQEVGTSYAVAVVRRSSHVT	120
Qy	121	IDTLKGVKSCHTGINTGNVNPVGVLYVESGRSLVNGCDVLKAVSDYFGGSCVPGAGETS	180
Db	121	IDTLKGVKSCHTGINTGNVNPVGVLYVESGRSLVNGCDVLKAVSDYFGGSCVPGAGETS	180
Qy	181	YSESRLCRGDSGSGGVCDKSPLERYDYSGAFRLAEGAGDVAFKHSTVLENTDGT	240
Db	181	YSESRLCRGDSGSGGVCDKSPLERYDYSGAFRLAEGAGDVAFKHSTVLENTDGT	240
Qy	241	LPSWGALLSQDFELLCRGSRADVTWEQCHLARVPAAHVAVVVRADTDGGLIFRLNNEGQ	300
Db	241	LPSWGALLSQDFELLCRGSRADVTWEQCHLARVPAAHVAVVVRADTDGGLIFRLNNEGQ	300
Qy	301	RLFSHSGSSQFMPSSSEAYGQKLLFKDSTSELVPIATQTYEAWLGHEYLHAMKGLCDPN	360
Db	301	RLFSHSGSSQFMPSSSEAYGQKLLFKDSTSELVPIATQTYEAWLGHEYLHAMKGLCDPN	360
Qy	361	RLPPYLRCVLTSTPEIKQCGMDMAVAFRRQLKPEIQCVSAKSPQHCMERIQAEVDATVL	420
Db	361	RLPPYLRCVLTSTPEIKQCGMDMAVAFRRQLKPEIQCVSAKSPQHCMERIQAEVDATVL	420

QY 421 SGEDIYTAGKKYGLVPAAGHVPEDSSNSYYVAVVRDSSHAFTLDELGRKRSCHAGF 480
 DB 421 SEDIYTAGKKYGLVPAAGHVPEDSSNSYYVAVVRDSSHAFTLDELGRKRSCHAGF 480
 QY 481 GSPAGWDVFPVGLIQRGFIIRPKDCDVLTAVERSEFFNASCVPVNNPKNYPSSLCALCVGDEQ 540
 DB 481 GSPAGWDVFPVGLIQRGFIIRPKDCDVLTAVERSEFFNASCVPVNNPKNYPSSLCALCVGDEQ 540
 QY 541 GRNKCVCNSQERYGYYGRGAFRCFLVENAGDVAFVRHTTVFDNTNGHNSPWAELRSEDYE 600
 DB 541 GRNKCVCNSQERYGYYGRGAFRCFLVENAGDVAFVRHTTVFDNTNGHNSPWAELRSEDYE 600
 QY 601 LLCPCNGARAEVSQFACNLIAQIIPPHAVMVRPDTNIFTVYGLDKAODLFGDDHNNKNGFKM 660
 DB 601 LLCPCNGARAEVSQFACNLIAQIIPPHAVMVRPDTNIFTVYGLDKAODLFGDDHNNKNGFKM 660
 QY 661 FDSSNYHGODLLFKDQATVRAVPVGEKTYRGMGLDYVAALBGMSSQCSGAAAPAPGAP 720
 DB 661 FDSSNYHGODLLFKDQATVRAVPVGEKTYRGMGLDYVAALBGMSSQCSGAAAPAPGAP 720
 QY 721 LLPALLPALAARLLPPAL 738
 DB 721 LLPALLPALAARLLPPAL 738

RESULT 2

US-08-520-933-3
 ; Sequence 3, Application US/08520933
 ; Patent No. 5981194
 ; GENERAL INFORMATION:
 ; APPLICANT: Jefferies, Wilfred A.
 ; APPLICANT: McGeer, Patrick L.
 ; APPLICANT: Rothenberger, Sylvia
 ; APPLICANT: Food, Michael R.
 ; APPLICANT: Yamada, Tatsuo
 ; APPLICANT: Kennard, Malcolm
 ; TITLE OF INVENTION: Use of p97 and Iron Binding Proteins
 ; TITLE OF INVENTION: as Diagnostic and Therapeutic Agents
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Bereskin & Parr
 ; STREET: 40 King Street West
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5H 3Y2
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/520,933
 ; FILING DATE: August 31, 1995
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Shona S. McDiarmid
 ; REGISTRATION NUMBER: 38,798
 ; REFERENCE/DOCKET NUMBER: 7685-006
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 416-364-7311
 ; TELEFAX: 416-361-1398
 ; TELEX: 06-23115
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 719 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-520-933-3

Query Match 97.6%; Score 3840; DB 2; Length 719;

Best Local Similarity 100.0%; Pred. No. 0;
 Matches 719; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 20 GMEVWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVOLIAAQEADAITLDG 79
 DB 1 GMEVWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVOLIAAQEADAITLDG 60
 QY 80 GAIYAGKEHGLKPVVGEVYDQEVGTSYVAVVRRSSHVITDILKGVKSCHTGINTVTG 139
 DB 61 GAIYAGKEHGLKPVVGEVYDQEVGTSYVAVVRRSSHVITDILKGVKSCHTGINTVTG 120
 QY 140 MNVPVGYLVESGRSLVSMGCDVLKAVSDYFGSCVPGAGETSYSSESLCRLCRGDSGEGVC 199
 DB 121 MNVPVGYLVESGRSLVSMGCDVLKAVSDYFGSCVPGAGETSYSSESLCRLCRGDSGEGVC 180
 QY 200 DKSPLEYYDYGAFRCFLAEGAGDVAFVKHSTVLENTDGTLPSSWGQALLSODFELLCRD 259
 DB 181 DKSPLEYYDYGAFRCFLAEGAGDVAFVKHSTVLENTDGTLPSSWGQALLSODFELLCRD 240
 QY 260 GSRADVTTEWROCHLARVPAHAVVVRADTDGGLIFRLNEGQRLFSHEGSSQFMSSEAYG 319
 DB 241 GSRADVTTEWROCHLARVPAHAVVVRADTDGGLIFRLNEGQRLFSHEGSSQFMSSEAYG 300
 QY 320 QKDLFLKDOTSELVPIATQTYEAWLGHYELHAMKGLLCPNRLPPYLKWCVLSTPEIQKC 379
 DB 301 QKDLFLKDOTSELVPIATQTYEAWLGHYELHAMKGLLCPNRLPPYLKWCVLSTPEIQKC 360
 QY 380 GDMVAFRRQRLKPEIQCVSAKSPQHMERIQAEQVDVATLSGEDIYTAGKKYGLVPAAG 439
 DB 361 GDMVAFRRQRLKPEIQCVSAKSPQHMERIQAEQVDVATLSGEDIYTAGKKYGLVPAAG 420
 QY 440 EHYAPEDSSNSYYVAVVRDSSHAFTLDELGRKRSCHAGFSPAGWDVFPVGLIQRGFI 499
 DB 421 EHYAPEDSSNSYYVAVVRDSSHAFTLDELGRKRSCHAGFSPAGWDVFPVGLIQRGFI 480
 QY 500 RPKDCDVLTAVERSEFFNASCVPVNNPKNYPSSI.CALCVGDEGRNKCVCNSQERYGYRGA 559
 DB 481 RPKDCDVLTAVERSEFFNASCVPVNNPKNYPSSI.CALCVGDEGRNKCVCNSQERYGYRGA 540
 QY 560 FRLCVENAGDVAFVRHTTVFDNTNGHNSPWAELRSEDYELLCPCNGARAEVSQFAACNL 619
 DB 541 FRLCVENAGDVAFVRHTTVFDNTNGHNSPWAELRSEDYELLCPCNGARAEVSQFAACNL 600
 QY 620 AQIPPHAVMVRPDTNIFTVYGLDKAODLFGDDHNNKNGFKMFDSSNYHGODLLFKDQATVR 679
 DB 601 AQIPPHAVMVRPDTNIFTVYGLDKAODLFGDDHNNKNGFKMFDSSNYHGODLLFKDQATVR 660
 QY 680 AVPVGEKTYRGMGLDYVAALBGMSSQCSGAAAPAPGAPLLPALLPALAARLLPPAL 738
 DB 661 AVPVGEKTYRGMGLDYVAALBGMSSQCSGAAAPAPGAPLLPALLPALAARLLPPAL 719

RESULT 3

US-09-285-040-3
 ; Sequence 3, Application US/09285040
 ; Patent No. 6455494
 ; GENERAL INFORMATION:
 ; APPLICANT: Jefferies, Wilfred A.
 ; APPLICANT: McGeer, Patrick L.
 ; APPLICANT: Rothenberger, Sylvia
 ; APPLICANT: Food, Michael R.
 ; APPLICANT: Yamada, Tatsuo
 ; APPLICANT: Kennard, Malcolm
 ; TITLE OF INVENTION: Use of p97 and Iron Binding Proteins
 ; TITLE OF INVENTION: as Diagnostic and Therapeutic Agents
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Bereskin & Parr
 ; STREET: 40 King Street West
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5H 3Y2

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/285,040
FILING DATE: 01-APR-1999
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gravelle, Micheline
REGISTRATION NUMBER: 40,261
REFERENCE/DOCKET NUMBER: 7685-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-364-7311
TELEFAX: 416-361-1398
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-285-040-3

Query Match 97.6%; Score 3840; DB 4; Length 719;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 719; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 20 GMEVRCATSDPEQHKCGNMSEAFREAGTQPSLLCVRGTSADHCVOLIAAQEADAITLDG 79
DB 1 GMEVRCATSDPEQHKCGNMSEAFREAGTQPSLLCVRGTSADHCVOLIAAQEADAITLDG 60
QY 80 GAIYEAGKEHGLKPVVGEVYDQEVGTSYAVAVVRRSSHVTIDTLGKVKSCHTGINRTVG 139
DB 61 GAIYEAGKEHGLKPVVGEVYDQEVGTSYAVAVVRRSSHVTIDTLGKVKSCHTGINRTVG 120
QY 140 MNVPVGYLVESGRSLVMGCDVLKAVSDYFGGSCVPGAGETSYSESLCRLCRGDSGEGVC 199
DB 121 MNVPVGYLVESGRSLVMGCDVLKAVSDYFGGSCVPGAGETSYSESLCRLCRGDSGEGVC 180
QY 200 DKSPLERYDYSGAFRLCAEGAGDVAFVKHSTVLENTDGTLPSSWQALLSQDFELLCRD 259
DB 181 DKSPLERYDYSGAFRLCAEGAGDVAFVKHSTVLENTDGTLPSSWQALLSQDFELLCRD 240
QY 260 GSRADVTWRQCHLARVPAHAVVVRADTDGGLIFRLINEGORLFSHEGSSFQMFSEAYG 319
DB 241 GSRADVTWRQCHLARVPAHAVVVRADTDGGLIFRLINEGORLFSHEGSSFQMFSEAYG 300
QY 320 QKDLLFKDSTSELVPIATQTYEAWLGHEYLHAWKGLLCDPNRLPPYLRCVLTSTPEIQKC 379
DB 301 QKDLLFKDSTSELVPIATQTYEAWLGHEYLHAWKGLLCDPNRLPPYLRCVLTSTPEIQKC 360
QY 380 GDMAVAFRRQRLKPEIQCVSAKSPQHCMERIOAEQVDVAVTLSGEDIYTAGKYGVLVPAAG 439
DB 361 GDMAVAFRRQRLKPEIQCVSAKSPQHCMERIOAEQVDVAVTLSGEDIYTAGKYGVLVPAAG 420
QY 440 EHYAPEDSSNSYYVAVVRDSSHAFTLDLRKRSCHAGFGSPAGWDVPVGGALIQRGFI 499
DB 421 EHYAPEDSSNSYYVAVVRDSSHAFTLDLRKRSCHAGFGSPAGWDVPVGGALIQRGFI 480
QY 500 RPKDCDVLTAVERSEFFNASCVPVNNPKNYPSSLCALCVGDEQGRKNCVGNSSQERYGYRGA 559
DB 481 RPKDCDVLTAVERSEFFNASCVPVNNPKNYPSSLCALCVGDEQGRKNCVGNSSQERYGYRGA 540
QY 560 FRCLVENAGDVAFVRHTTTFDNTNGHNSPWAELRSEDEYELLCNPGARAEVSQFAACNL 619
DB 541 FRCLVENAGDVAFVRHTTTFDNTNGHNSPWAELRSEDEYELLCNPGARAEVSQFAACNL 600
QY 620 AOIPPHAVVVRPDNTNIFTVYGLLDKAQDLFGDDHKNKGKFMFDSNHYHGODLLFKDQATVR 679
DB 601 AOIPPHAVVVRPDNTNIFTVYGLLDKAQDLFGDDHKNKGKFMFDSNHYHGODLLFKDQATVR 660
QY 680 AVPVGEKTTYRGWGLDYVAALLEGMSQQCSGAAAPAPGAPLPLLLPALAARLLPPAL 738
```

```
DB 661 AVPVGEKTTYRGWGLDYVAALLEGMSQQCSGAAAPAPGAPLPLLLPALAARLLPPAL 719
RESULT 4
5262177-5
; Patent No. 5262177
; APPLICANT: BROWN, J OSEPH P.; ESTIN, CHARLES D.; PLOWMAN, GREGORY
; D. HELLSTROM, KARL E.; ROSE, TIMOTHY M.; HELLSTROM, INGEGRD;
; PURCHIO, ANTHONY F.; HU, SHIU-LOK; PENNATHUR, SRIDHAR
; TITLE OF INVENTION: RECOMBINANT VIRUSES ENCODING THE HUMAN
; MELANOMA-ASSOCIATED ANTIGEN
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/7230
; FILING DATE: 27-JAN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 827,313
; FILING DATE: 07-FEB-1986
; SEQ ID NO: 5
; LENGTH: 717
5262177-5
```

Query Match 96.3%; Score 3792; DB 6; Length 717;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 714; Conservative 1; Mismatches 2; Indels 2; Gaps 2;

```
QY 20 GMEVRCATSDPEQHKCGNMSEAFREAGTQPSLLCVRGTSADHCVOLIAAQEADAITLDG 79
DB 1 GMEVRCATSDPEQHKCGNMSEAFREAGTQPSLLCVRGTSADHCVOLIAAQEADAITLDG 60
QY 80 GAIYEAGKEHGLKPVVGEVYDQEVGTSYAVAVVRRSSHVTIDTLGKVKSCHTGINRTVG 139
DB 61 GAIYEAGKEHGLKPVVGEVYDQEVGTSYAVAVVRRSSHVTIDTLGKVKSCHTGINRTVG 120
QY 140 MNVPVGYLVESGRSLVMGCDVLKAVSDYFGGSCVPGAGETSYSESLCRLCRGDSGEGVC 199
DB 121 MNVPVGYLVESGRSLVMGCDVLKAVSDYFGGSCVPGAGETSYSESLCRLCRGDSGEGVC 180
QY 200 DKSPLERYDYSGAFRLCAEGAGDVAFVKHSTVLENTDGTLPSSWQALLSQDFELLCRD 259
DB 181 DKSPLERYDYSGA -KCLAEAGAGDVAFVKHSTVLENTDGTLPSSWQALLSQDFELLCRD 239
QY 260 GSRADVTWRQCHLARVPAHAVVVRADTDGGLIFRLINEGORLFSHEGSSFQMFSEAYG 319
DB 240 GSRADVTWRQCHLARVPAHAVVVRADTDGGLIFRLINEGORLFSHEGSSFQMF -FEAYG 298
QY 320 QKDLLFKDSTSELVPIATQTYEAWLGHEYLHAWKGLLCDPNRLPPYLRCVLTSTPEIQKC 379
DB 299 QKDLLFKDSTSELVPIATQTYEAWLGHEYLHAWKGLLCDPNRLPPYLRCVLTSTPEIQKC 358
QY 380 GDMAVAFRRQRLKPEIQCVSAKSPQHCMERIOAEQVDVAVTLSGEDIYTAGKYGVLVPAAG 439
DB 359 GDMAVAFRRQRLKPEIQCVSAKSPQHCMERIOAEQVDVAVTLSGEDIYTAGKYGVLVPAAG 418
QY 440 EHYAPEDSSNSYYVAVVRDSSHAFTLDLRKRSCHAGFGSPAGWDVPVGGALIQRGFI 499
DB 419 EHYAPEDSSNSYYVAVVRDSSHAFTLDLRKRSCHAGFGSPAGWDVPVGGALIQRGFI 478
QY 500 RPKDCDVLTAVERSEFFNASCVPVNNPKNYPSSLCALCVGDEQGRKNCVGNSSQERYGYRGA 559
DB 479 RPKDCDVLTAVERSEFFNASCVPVNNPKNYPSSLCALCVGDEQGRKNCVGNSSQERYGYRGA 538
QY 560 FRCLVENAGDVAFVRHTTTFDNTNGHNSPWAELRSEDEYELLCNPGARAEVSQFAACNL 619
DB 539 FRCLVENAGDVAFVRHTTTFDNTNGHNSPWAELRSEDEYELLCNPGARAEVSQFAACNL 598
QY 620 AOIPPHAVVVRPDNTNIFTVYGLLDKAQDLFGDDHKNKGKFMFDSNHYHGODLLFKDQATVR 679
DB 599 AOIPPHAVVVRPDNTNIFTVYGLLDKAQDLFGDDHKNKGKFMFDSNHYHGODLLFKDQATVR 658
QY 680 AVPVGEKTTYRGWGLDYVAALLEGMSQQCSGAAAPAPGAPLPLLLPALAARLLPPAL 738
```

Db 659 AVPEGKTTYRGLGIDYVAALGKMSQSCGAAPAPGAPLLPLLLPALAALLPPAL 717

RESULT 5

US-08-459-818-19
; Sequence 19, Application US/08459818
; Patent No. 5851795

GENERAL INFORMATION:

APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damle, Nitin K.
APPLICANT: Brady, William
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:

ADDRESSES: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FastSeq 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,818
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436.35US02
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 502 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-459-818-19

Query Match 47.0%; Score 1850; DB 2; Length 502;

Best Local Similarity 100.0%; Pred. No. 3.1e-172;

Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	20	GMEVWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVQLIAAQEADAITLDG	79
Db	156	GMEVWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVQLIAAQEADAITLDG	215
QY	80	GAIEAGKEHGLKPVVGEVYDQVGTSTYYAVAVRRSSHVTIDTLKGVKSCHTGINRTVG	139
Db	216	GAIEAGKEHGLKPVVGEVYDQVGTSTYYAVAVRRSSHVTIDTLKGVKSCHTGINRTVG	275
QY	140	WNVPVGLVSGRLSVMGCDVLKAVSDYFGGSCVPGAGETSYESLCLRCRGSSEGGVC	199
Db	276	WNVPVGLVSGRLSVMGCDVLKAVSDYFGGSCVPGAGETSYESLCLRCRGSSEGGVC	335
QY	200	DKSPLEYYDYGAFRCLEAGAGDVA FVKHSTVLENTDGTLPNMQALLSQDFELLCRD	259
Db	336	DKSPLEYYDYGAFRCLEAGAGDVA FVKHSTVLENTDGTLPNMQALLSQDFELLCRD	395
QY	260	GSRADVTWROCHLARVPAHVVVRADTDGGLIFRLNEGQRLFSHGSSFQMFSSAYG	319
Db	396	GSRADVTWROCHLARVPAHVVVRADTDGGLIFRLNEGQRLFSHGSSFQMFSSAYG	455
QY	320	QKDLFFKDSSTSELVPIATQTYEAWLGHYHAKMGLLCDPNRLPPYL	366
Db	456	QKDLFFKDSSTSELVPIATQTYEAWLGHYHAKMGLLCDPNRLPPYL	502

RESULT 6

US-08-889-666-19
; Sequence 19, Application US/08889666
; Patent No. 5885579

GENERAL INFORMATION:

APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damle, Nitin K.
APPLICANT: Brady, William
APPLICANT: Kiener, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:

ADDRESSES: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/889,666
FILING DATE: 08-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-35US01
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 502 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-889-666-19

Query Match 47.0%; Score 1850; DB 2; Length 502;

Best Local Similarity 100.0%; Pred. No. 3.1e-172;

Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	20	GMEVWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVQLIAAQEADAITLDG	79
Db	156	GMEVWCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVQLIAAQEADAITLDG	215
QY	80	GAIEAGKEHGLKPVVGEVYDQVGTSTYYAVAVRRSSHVTIDTLKGVKSCHTGINRTVG	139
Db	216	GAIEAGKEHGLKPVVGEVYDQVGTSTYYAVAVRRSSHVTIDTLKGVKSCHTGINRTVG	275
QY	140	WNVPVGLVSGRLSVMGCDVLKAVSDYFGGSCVPGAGETSYESLCLRCRGSSEGGVC	199
Db	276	WNVPVGLVSGRLSVMGCDVLKAVSDYFGGSCVPGAGETSYESLCLRCRGSSEGGVC	335
QY	200	DKSPLEYYDYGAFRCLEAGAGDVA FVKHSTVLENTDGTLPNMQALLSQDFELLCRD	259
Db	336	DKSPLEYYDYGAFRCLEAGAGDVA FVKHSTVLENTDGTLPNMQALLSQDFELLCRD	395
QY	260	GSRADVTWROCHLARVPAHVVVRADTDGGLIFRLNEGQRLFSHGSSFQMFSSAYG	319
Db	396	GSRADVTWROCHLARVPAHVVVRADTDGGLIFRLNEGQRLFSHGSSFQMFSSAYG	455

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

200 DKSPLERYDYDVGAFRLAEGADVAFVGHSTVLENTDGTLLPSNGOALLSQDFELLCRD 259

Db 336 DKSPLEYYDYGAFRCLEAGAGDVAFVKHSTVLENTDGTLPWSGQALLSQDFELLCRD 395
QY 260 GSRADVTEWROCHLARVPAHVVVRADTDGGLIFRLNNEGQRLFSHEGSSFQMFSEAYG 319
Db 396 GSRADVTEWROCHLARVPAHVVVRADTDGGLIFRLNNEGQRLFSHEGSSFQMFSEAYG 455
QY 320 QKDLFFKDSSTSELVPIATQTYEAWLGHEYLHAKMGLLCDPNRLPPYL 366
Db 456 QKDLFFKDSSTSELVPIATQTYEAWLGHEYLHAKMGLLCDPNRLPPYL 502

RESULT 9

US-08-488-062-19
; Sequence 19, Application US/08488062
; Patent No. 5977318
; GENERAL INFORMATION:
; APPLICANT: Linsley, Peter S.
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Damle, Nitin K.
; APPLICANT: Brady, William
; APPLICANT: Kienet, Peter A.
; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 11150 Santa Monica Blvd., Suite 400
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,062
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/375390
; FILING DATE: 18-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Adriano, Sarah B.
; REGISTRATION NUMBER: 34,470
; REFERENCE/DOCKET NUMBER: 30436-35US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310-445-1140
; TELEFAX: 310-445-9031
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 502 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-488-062-19

Query Match 47.0%; Score 1850; DB 2; Length 502;
Best Local Similarity 100.0%; Pred. No. 3.1e-172;
Matches 347; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 GNEVRCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVOLIAAQADAITLDG 79
Db 156 GNEVRCATSDPEQHKCGNMSEAFREAGIQPSLLCVRGTSADHCVOLIAAQADAITLDG 215
QY 80 GAIEYAGKEHGLKPVVGEYVDQEVGYSYAVAVVRSSHVTIDTLKGKVSCHTGNRTVG 139
Db 216 GAIEYAGKEHGLKPVVGEYVDQEVGYSYAVAVVRSSHVTIDTLKGKVSCHTGNRTVG 275
QY 140 WNPVGYLVESGRLSVMGCDVLKAVSDYFEGGSCVPGAGETSYSESICRLCRGDSGEGVC 199

Db 276 WNPVGYLVESGRLSVMGCDVLKAVSDYFEGGSCVPGAGETSYSESICRLCRGDSGEGVC 335
QY 200 DKSPLEYYDYGAFRCLEAGAGDVAFVKHSTVLENTDGTLPWSGQALLSQDFELLCRD 259
Db 336 DKSPLEYYDYGAFRCLEAGAGDVAFVKHSTVLENTDGTLPWSGQALLSQDFELLCRD 395
QY 260 GSRADVTEWROCHLARVPAHVVVRADTDGGLIFRLNNEGQRLFSHEGSSFQMFSEAYG 319
Db 396 GSRADVTEWROCHLARVPAHVVVRADTDGGLIFRLNNEGQRLFSHEGSSFQMFSEAYG 455
QY 320 QKDLFFKDSSTSELVPIATQTYEAWLGHEYLHAKMGLLCDPNRLPPYL 366
Db 456 QKDLFFKDSSTSELVPIATQTYEAWLGHEYLHAKMGLLCDPNRLPPYL 502

RESULT 10

US-08-175-158A-2
; Sequence 2, Application US/08175158A
; Patent No. 5986067
; GENERAL INFORMATION:
; APPLICANT: FUNK, Walter D.
; APPLICANT: MacGILLIVRAY, Ross T.A.
; APPLICANT: MASON, Anne B.
; APPLICANT: WOODWORTH, Robert C.
; TITLE OF INVENTION: RECOMBINANT TRANSFERRINS, TRANSFERRIN HALF-
; TITLE OF INVENTION: MOLECULES AND MUTANTS THEREOF
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/175,158A
; FILING DATE: 28-DEC-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/832,029
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DeConti, Giulio A.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: UVI-005CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 698 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-175-158A-2

Query Match 32.3%; Score 1269.5; DB 2; Length 698;
Best Local Similarity 40.4%; Pred. No. 3.9e-115;
Matches 304; Conservative 118; Mismatches 221; Indels 109; Gaps 32;
QY 1 MRPGSGALWL--LLALRTVLGMEYRWCATSDPEQHKCGNMSEAFRE-----AGIQPSL 52
Db 1 MRLAVGALLVCVGLGCLAVPKTVRCVASEHEATKC---QSFDRHMKSVIPSDGPSV 56
QY 53 LCVRTGTSADHCVOLIAAQADAITLDGGAIVYAG-KEHGLKPVVGEVY--DOEVGTSYYA 109
Db 57 ACVKKASYLDICITRAANNEADAVTLDAVLVDYLAIPNNLKPWVAEFYGSKEPDPTFYA 116

```

QY 110 VAVVRRSHVITDITLKGKVSCHTGINRTVGMNVPVGVYVSGRLSVMGCD-----VLK 162
Db 117 VAVVKDQSGMGNQRLKSKCHTGLGSGAGNIPIGLLY-----CDLPEPRKPLEK 167
QY 163 AVSDYFGGSCVPAGNETSYSESICRLCRGSSGEGVCDKSPLERYDYSGAFRCLEAGAG 222
Db 168 AVANFSGSCAPAGDGTDFPO-LCQLCPG-----CGCSTLNQYFGYSGAFKCLDAG 219
QY 223 DVAFVXKSTVLENTDGTKLPMSGQALLSOD-FELLCRDGSRADVTWROCHLARVPAHAV 281
Db 220 DVAFVXKSTIFENLANKA-----DRQYELLCLDNTKPKVDEYKDCHLAQVPSHTV 270
QY 282 VVRADTDG--LIFRLNQGRLFSHEGS-SFQMFSSSEAYGQKDLIFKDDSTSELVPIAT 337
Db 271 VAR--SMGKEDLIWELLNQAQHFQKDKSKFQLFSS-PHG-KDLLFKDSAHGFLKVP 326
QY 338 Q-TVEAWLGHXYLHAKGLL-----CDPNRLPPYLRCVLSSTPIQKCGDMAVA 385
Db 327 RMDAKMYLGYEYTAIRNREGTCPEAPTDECKP-----VKWALSHERLKCDEWSV- 379
QY 386 FRQRLKPEIQCVSAKSPOHMERIOAQVDAVTLGSDIYTAGKYGLVPAAGHYAPE 445
Db 380 ---NSVGKIECVSAETTEDCIAKIMNGEADMSLDGFFYIAG-KCGLVPLVLAENYKS 434
QY 446 DSSN-----SYVAVVVRDSSHAFTLDELGRKSGCHAGFGSPAGHDVPVGCALIQGFIR 500
Db 435 DNCDTPEAGYFAVAVVVKASD-LTWDLNKGKSKCHTAVGRTAGNIPMGLL-----YNK 489
QY 501 PKDCDVLTVSEFFNASCVPVNNPKNYPSSILCALCVGDEGRNKCNGNSQERYGYRGAP 560
Db 490 INHC-----RDEFFSEGCAGSKK---DSSLCKLCMG--SGLNLCEPNNKEGYGTGAF 540
QY 561 RCLVENAGDVAFVRHTTVFNTNGHNSPEWAAELRSEDYELLCPNGARAEVSQFAACNLA 620
Db 541 RCLVEK-GDVAFKVQTPQNTGKPNPDKWAKNLEKDYELLCLDGTGRKEVEYANCHLA 599
QY 621 QIPPHAVMVPDNNITFVYGLLQKADLFGDDHN--KNGKFMFDSSNYHGQDLFFKDATV 678
Db 600 RAPNHAHVTRKDEA-CVHKILRQOQHLFGSNVTDGSGNFCLFRSET---KDLLFRDVT 655
QY 679 RAVPVKEKTYRGWGLGDVVAALGEMSSQOCS 710
Db 656 CLAKLHNRNTEYKLGEEYKAVGNL--RKCS 685

```

RESULT 11

```

US-08-470-058-2
; Sequence 2, Application US/08470058
; Patent No. 5817789
; GENERAL INFORMATION:
; APPLICANT: Heartlein, Michael W.
; APPLICANT: Lemontt, Jeffrey P.
; TITLE OF INVENTION: Chimeric Proteins For Use in Transport
; TITLE OF INVENTION: of a Selected Substance Into Cells
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,058
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia

```

```

; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: TKT93-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1074 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-470-058-2

Query Match 32.2%; Score 1268.5; DB 2; Length 1074;
Best Local Similarity 40.5%; Pred. No. 1e-114;
Matches 299; Conservative 117; Mismatches 215; Indels 107; Gaps 31;

QY 13 ALRTVLGGMEVRWCATSDPEQHKCGNMSEAFRE-----AGIQPSLLCVRGTSADHCVOL 66
Db 391 ACKAVVPDKTVRWCAVSEHEATKC-----QSFDRHMKSVIPSDGSPVACVKKASYLDCIRA 446
QY 67 IAAQEAADAITLDGGAITYEAG-KEHGLKPVVGEVY--DOEVGTSYYAVAVVRRSSHVTIDT 123
Db 447 IAAEADAVTLDAGLVYDAYLAPNLLKPVVAEFGSKEDPQTFYAVAVVVKDQSGFQMNO 506
QY 124 LKGVSCHTGGINRTVGMNVPVGVYVSGRLSVMGCD-----VLKAVSDYFGGSCVPGA 176
Db 507 LRGKKSCHTGLGSGAGNIPIGLLY-----CDLPEPRKPLEKAVANFFSGSCAPCA 557
QY 177 GETSYSESICRLCRGSSGEGVCDKSPLERYDYSGAFRCLEAGADGVAFVKHSTVLENT 236
Db 558 DGTDFPO-LCQLCPG-----CGCSTLNQYFGYSGAFKCLDAGDAGVAFVKHSTIFENL 609
QY 237 DGTLPSPWQALLSOD-FELLCRDGSRADVTWROCHLARVPAHAVVVRADTDG--LI 292
Db 610 ANKA-----DRQYELLCLDNTKPKVDEYKDCHLAQVPSHTVVAR--SMGKEDLI 658
QY 293 FRLLNQGRLFSHEGS-SFQMFSSSEAYGQKDLIFKDDSTSELVPIATQ-TYEAWLGHXYLH 350
Db 659 WELLNQAQHFQKDKSKFQLFSS-PHG-KDLLFKDSAHGFLKVPFRMDAKMYLGYEYVT 716
QY 351 AMKGLL-----CDPNRLPPYLRCVLSSTPIQKCGDMAVAFRRQRLKPEIQCVS 399
Db 717 AIRNLREGTCPEAPTDECKP-----VKWALSHERLKCDEWSV-----NSVGKIECVS 765
QY 400 AKSPQHMERIOAQVDAVTLGSDIYTAGKYGLVPAAGHYAPEDSSN-----SYVYV 454
Db 766 AETTEDCIAKIMNGEADMSLDGFFYIAG-KCGLVPLVLAENYKNSDNCEDTPEAGYFAV 824
QY 455 AVVRDSSHAFTLDELGRKSGCHAGFGSPAGWDVPVGCALIQGFIRPKDCDVLTVASEFF 514
Db 825 AVVKASD-LTWDLNKGKSKCHTAVGRTAGNIPMGLL-----YKINHC-----RDEFF 875
QY 515 NASCVFVNNPKNYPSSILCALCVGDEGRNKCNGNSQERYGYRGAFRCLEVENAGDVAFVR 574
Db 876 SEGCAFGSKK---DSSLCKLCMG--SGLNLCEPNNKEGYGTGAFRCLEK-GDVAFVK 929
QY 575 HTTVFNTNGHNSPEWAAELRSEDYELLCPNGARAEVSQFAACNLAQIPPHAVMVPDNN 634
Db 930 HQTVPQNTGKPNPDKWAKNLEKDYELLCLDGTGRKEVEYANCHLARAPNHAHVTRKDE 989
QY 635 IFTVYGLLDKADLFGDDHN--KNGKFMFDSSNYHGQDLFFKDATVRAVPVVGKTTYRGW 692
Db 990 A-CVHKILRQOQHLFGSNVTDGSGNFCLFRSET---KDLLFRDVTCLAKLHNRNTEYKY 1045
QY 693 LGLDYVAALGEMSSQOCS 710
Db 1046 LGEEYKAVGNL--RKCS 1061

```

RESULT 12

US-09-037-188-2

Query Match 32.2%; Score 1268.5; DB 3; Length 1074;

QY 645 AQDLFGDDHN--KNGFKMFDSSNYHGQDLFLFKDATVRAVPVGEKTTTGRWGLGLDYVAALE 702
 Db 1335 QQHFLGSGNVTDCSGNFCFLRSET---KDLFRDDTVCLAKLHNRNTYKYLGEYVKAAG 1391
 QY 703 GMSOQCS 710
 Db 1392 NL--RKCS 1397

RESULT 15

US-09-037-188-4
 ; Sequence 4, Application US/09037188
 ; Patent No. 6027921
 ; GENERAL INFORMATION:
 ; APPLICANT: Heartlein, Michael W.
 ; APPLICANT: Lemontt, Jeffrey F.
 ; APPLICANT: Concino, Michael F.
 ; TITLE OF INVENTION: CHIMERIC PROTEINS FOR USE
 ; TITLE OF INVENTION: IN TRANSPORT OF A SELECTED SUBSTANCE INTO CELLS
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: Windows95
 ; SOFTWARE: FastSeq for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/037,188
 ; FILING DATE: 02-MAR-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE: 06-JUN-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fraser, Ph.D., J.D., Janis K.
 ; REGISTRATION NUMBER: 34,919
 ; REFERENCE/DOCKET NUMBER: 07236/009002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617/542-5070
 ; TELEFAX: 617/542-8906
 ; TELEX: 200154
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1410 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; US-09-037-188-4

Query Match 32.1%; Score 1263.5; DB 3; Length 1410;
 Best Local Similarity 40.8%; Pred. No. 4.9e-114;
 Matches 297; Conservative 115; Mismatches 209; Indels 107; Gaps 31;
 QY 23 VRWCATSDPEQHKCGNMSEAFRE-----AGIQPSLLCYRGTSADHCVQLIAAQEADAIT 76
 Db 737 VRWCATSEHATKC-----QSFDRHMKSVIPSDGPSVACVKKASYLDCIRAIANEADAVT 792
 QY 77 LDGAIYEAG-KEHGLKPVVGEVY--DQEVGTSYAVAVVRRSSHVTIDTLKGVKSCHTG 133
 Db 793 LDAGLVYDAYLAPNNLKPVABFYGSKEDPQTFYAVAVVKKDSGFQMNQLRGKSCHTG 852
 QY 134 INRTVGMNVPGVYVESGRSLVWGC-----VLKAVSDYFGGSCVPGAGETSYSESIC 186
 Db 853 LGRSAGWNPILGLY-----CDLPEPRKPLEKAVANFFSGSCAPCADGTDFFQ-LC 902
 QY 187 RLCRGDSGEGVCDKSPLERYDYSGAFCLAEAGADVAFVXHSVTLENTDKTLPSWQ 246
 Db 903 QLCPG-----CGCSTLNQYFGYSGAFKCLKDGAGDVAFVXHSVTIFENLANKA----- 949

Search completed: May 14, 2004, 09:44:34
 Job time : 26 secs

QY 247 ALLSOD-FELLCRGSRADVTETWROCHLARVPAPAHAVVVRADTDGG---LIFRLNEGQRL 302
 Db 950 ---DRDQVELLCLDNTRKPVDEYKDLHAQVPSHTVVAR--SMGSKEDLIWELLNQAEH 1004
 QY 303 FSHEGS-SFQMFSEAYGQKDLLFKDSTSELVPIATQ-TYEAWLGHGHEYLHAKMGLL---- 356
 Db 1005 FGKDKSKEFQLFSS-PHG-KDLLPKDSAHGFLKVPPRMDAKWYLGVEYVVTAINRLEGTG 1062
 QY 357 -----CDPNRLPPYLRWCVLSTPEIOKCGDMAVAFRRQRILKPEIOCVSAKSPHOCMER 409
 Db 1063 PEAPTDECKP-----VKWCALSHHERLKCDWSV-----NSVGKIECVSAETTEDCIAK 1111
 QY 410 IQAEQVDATVLSGEDIYTAGKYGVLVPAAGHYAPEDSSN-----SYVVAVVRDSSHA 464
 Db 1112 IMNGEADAMSLDGGFVYIAG-KCGLVPVLAENYKSDNCEDTPEAGYFAVAVVKASD- 1169
 QY 465 FTLDLGRKRSCHAGFGSPAGWDVPVGALIQRGFIRPKDCDVLTAVERSEFFNASCVPVNNP 524
 Db 1170 LTWDNLKGGKCHTAGVGTAGWNIPMGLL----YKINHC-----RFDEFFSEGCAPGSKK 1221
 QY 525 KNPSSLCALCVGDEQGRNKCNGSQERYGYRGAFRCCLVENAGDVAFVRHTTTFVDNTNG 584
 Db 1222 ---DSSLCKLWCG--SGLNLCEPNNKEGYGYGTGAFRCCLVEK-GDVAFAVKGHTVPQNTGG 1275
 QY 585 HNSEPAAELRSEDYELLCPNGARAEVSQFAACNLAQIPPHAVMVRPDTNIFTVYGLLDK 644
 Db 1276 KNPDPAKNLNEKYELLCLDGTGTRKPVVEYANCHLAPNHAHVTRKDKEA-CVHKILRQ 1334
 QY 645 AQDLFGDDHN--KNGFKMFDSSNYHGQDLFLFKDATVRAVPVGEKTTTGRWGLGLDYVAALE 702
 Db 1335 QQHFLGSGNVTDCSGNFCFLRSET---KDLFRDDTVCLAKLHNRNTYKYLGEYVKAAG 1391
 QY 703 GMSOQCS 710
 Db 1392 NL--RKCS 1397